



**HAL**  
open science

# Modèle de régression pour des données non-Euclidiennes en grande dimension. Application à la classification de taxons en anatomie computationnelle.

Thi Thien Trang Bui

## ► To cite this version:

Thi Thien Trang Bui. Modèle de régression pour des données non-Euclidiennes en grande dimension. Application à la classification de taxons en anatomie computationnelle.. Mathématiques générales [math.GM]. INSA de Toulouse, 2019. Français. NNT : 2019ISAT0021 . tel-02497588

**HAL Id: tel-02497588**

**<https://theses.hal.science/tel-02497588v1>**

Submitted on 3 Mar 2020

**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.

# Contents

<b>Résumé</b>	<b>i</b>
<b>Abstract</b>	<b>i</b>
<b>1 Introduction générale</b>	<b>11</b>
1.1 Motivation . . . . .	11
1.1.1 Le modèle de régression . . . . .	12
1.1.2 Test d’hypothèses en régression non paramétrique . . . . .	13
1.2 Objectifs . . . . .	14
1.3 Présentation des résultats . . . . .	16
1.3.1 Régression dans un RKHS pour des distributions: application à la prévision de l’age en fonction des courbes d’oto-émission . . . . .	16
1.3.2 Tests sur courbes . . . . .	20
<b>I Part 1: Gaussian kernel regression</b>	<b>31</b>
<b>2 Distribution regression model with a Reproducing Kernel Hilbert Space approach</b>	<b>33</b>
2.1 Introduction . . . . .	34
2.2 Preliminaries . . . . .	35
2.2.1 Wasserstein space and barycenter of a probability in Wasser- stein space . . . . .	35
2.2.2 Reproducing kernel Hilbert space . . . . .	37
2.3 Kernel on Wasserstein space $\mathcal{W}_2(\mathbb{R})$ . . . . .	38
2.3.1 The Wasserstein space on $\mathbb{R}$ . . . . .	38
2.3.2 Kernel . . . . .	40
2.4 Regression . . . . .	41
2.4.1 Setting . . . . .	41
2.4.2 Universal kernel . . . . .	42

2.5	Numerical Simulations and Real data application . . . . .	46
2.5.1	Simulation in the regression model with one-dimensional distribution input . . . . .	46
2.5.2	Simulation in the regression model with two-dimensional distribution input . . . . .	52
2.5.3	Application on evolution of hearing sensitivity . . . . .	55
2.6	Discussion . . . . .	59
 <b>II Part 2: Testing hypothesis</b>		<b>61</b>
<b>3</b>	<b>Aggregated kernel based tests for signal detection in a regression model</b>	<b>63</b>
3.1	Introduction . . . . .	64
3.2	Single tests based on a single kernel. . . . .	66
3.2.1	Definition of the testing procedure. . . . .	66
3.2.2	Probabilities of first and second kind errors of the test. . . . .	68
3.2.3	Performance of the Monte Carlo approximation. . . . .	70
3.3	Two particular examples of kernel function. . . . .	71
3.3.1	Projection kernels. . . . .	71
3.3.2	Gaussian kernels. . . . .	73
3.4	Multiple or aggregated tests based on collections of kernel functions. . . . .	74
3.4.1	The aggregated testing procedure. . . . .	75
3.4.2	The aggregation of projection kernels. . . . .	75
3.4.3	The aggregation of Gaussian kernels. . . . .	77
3.5	Simulation study. . . . .	78
3.5.1	Presentation of the simulation study. . . . .	78
3.5.2	Simulation results. . . . .	79
3.6	Appendices . . . . .	81
3.6.1	Proof of Proposition 3.1 . . . . .	81
3.6.2	Proof of Theorem 3.2 . . . . .	86
3.6.3	Proof of Corollary 3.5 and 3.7. . . . .	87
3.6.4	Proof of Proposition 3.6 . . . . .	87
3.6.5	Proof of Proposition 3.8 . . . . .	88
3.6.6	Proof of Theorem 3.9, Corollary 3.10 and 3.12. . . . .	89
3.6.7	Proof of Corollary 3.11. . . . .	90
3.6.8	Proof of Corollary 3.12. . . . .	90

---

**4 Conclusion et perspectives****93**



# Résumé

Dans cette thèse, nous étudions un modèle de régression avec des entrées de type distribution et le problème de test d'hypothèse pour la détection de signaux dans un modèle de régression. Nos modèles ont été appliqués aux données de sensibilité auditive mesurées par otoémissions acoustiques <sup>1</sup>, cette mesure biologique contenant potentiellement des informations annexes sur l'individu (age, sexe, population/espèce).

Dans la première partie, un nouveau modèle de régression de distribution pour les distributions de probabilité est introduit. Ce modèle est basé sur un cadre de régression RKHS (Reproducing Kernel Hilbert Space), dans lequel les noyaux universels sont construits à l'aide de distances de Wasserstein pour les distributions appartenant à  $\mathcal{W}_2(\Omega)$  où  $\Omega$  est un sous-espace compact de  $\mathbb{R}$ . Nous prouvons la propriété de noyau universel de ces noyaux et utilisons ce cadre pour effectuer des régressions sur des fonctions. Différents modèles de régression sont d'abord comparés à celui proposé sur des données fonctionnelles simulées. Nous appliquons ensuite notre modèle de régression aux réponses de distribution des émissions otoacoustiques évoquées transitoires (TEOAE) et aux prédicteurs réels de l'âge. Cette partie est un travail en collaboration avec Loubes, J-M., Risser, L. et Balaesque, P..

Dans la deuxième partie, en considérant un modèle de régression, nous abordons la question du test de la nullité de la fonction de régression. Nous proposons tout d'abord une nouvelle procédure de test unique basée sur un noyau symétrique général et une estimation de la variance des observations. Les valeurs critiques correspondantes sont construites pour obtenir des tests non-asymptotiques de niveau  $\alpha$ . Nous introduisons ensuite une procédure d'agrégation afin d'éviter le choix complexe du noyau et des paramètres de celui-ci. Les tests multiples vérifient les propriétés non asymptotiques et adaptatives au sens minimax sur plusieurs classes d'alternatives régulières.

**Mots-clés:** Régression, reproduction de l'espace de Hilbert du noyau, distance de Wasserstein, émission otoacoustique évoquée transitoire, taux de séparation, tests adaptatifs, méthodes du noyau, test agrégé.

---

<sup>1</sup>vibrations de nature sonore générées par les mouvements des cellules ciliées externes, situées le long de la membrane basilaire cochléaire - Wikipédia.



# Abstract

In this thesis, we study a regression model with distribution entries and the testing hypothesis problem for signal detection in a regression model. We aim to apply these models in hearing sensitivity measured by the transient evoked otoacoustic emissions (TEOAEs) data to improve our knowledge in the auditory investigation.

In the first part, a new distribution regression model for probability distributions is introduced. This model is based on a Reproducing Kernel Hilbert Space (RKHS) regression framework, where universal kernels are built using Wasserstein distances for distributions belonging to  $\mathcal{W}_2(\Omega)$  and  $\Omega$  is a compact subspace of  $\mathbb{R}$ . We prove the universal kernel property of such kernels and use this setting to perform regressions on functions. Different regression models are first compared with the proposed one on simulated functional data. We then apply our regression model to transient evoked otoacoustic emission (TEOAE) distribution responses and real predictors of the age. This part is a joint work with Loubes, J-M., Risser, L. and Balaresque, P..

In the second part, considering a regression model, we address the question of testing the nullity of the regression function. The testing procedure is available when the variance of the observations is unknown and does not depend on any prior information on the alternative. We first propose a single testing procedure based on a general symmetric kernel and an estimation of the variance of the observations. The corresponding critical values are constructed to obtain non asymptotic level- $\alpha$  tests. We then introduce an aggregation procedure to avoid the difficult choice of the kernel and of the parameters of the kernel. The multiple tests satisfy non-asymptotic properties and are adaptive in the minimax sense over several classes of regular alternatives.

**Keywords:** Regression, Reproducing kernel Hilbert space, Wasserstein distance, Transient evoked otoacoustic emission, separation rates, adaptive tests, kernel methods, aggregated test.





# Remerciements

J'adresserais en premier lieu mes sincères remerciements à mes directeurs de thèse Jean-Michel Loubes et Patricia Balaesque de m'avoir permis de travailler dans le milieu de la recherche pendant trois ans. Ils ont passé beaucoup de temps à s'occuper de moi, à m'encourager, à partager toutes leurs idées sur mon sujet de thèse. Merci beaucoup Jean Michel de me inspirer d'examiner toutes les questions pratiques avec le point de vue d'un statisticien. Merci beaucoup Patricia pour votre enthousiasme à m'aider les derniers mois, c'est particulièrement significatif pour moi. J'ai trouvé la passion de travailler avec des données réelles, en particulier avec des données cochléaires, cela a orienté ma future carrière plus clairement.

J'exprime aussi mes plus sincères remerciements à Béatrice Laurent Bonneau, Xavier Bressau et Dang Duc Trong qui m'ont aidé à obtenir une bourse de Master 2 en Mathématiques Fondamentales et Applications à Toulouse. Ceci marquera le début de mes études en France. Je remercie plus particulièrement Béatrice de m'avoir aidée à mon arrivée en France et avoir beaucoup m'avoir transmise toute son expérience. Je n'oublierai jamais l'instant quand elle m'a donné une couverture et une écharpe, c'était un moment rempli d'émotion pour une étudiante qui viens d'arrivé dans un pays à quelques milliers de kilomètres de sa famille. Je remercie tout aussi particulièrement Trong pour ses mots d'encouragement et ses conseils pendant cinq ans, il est comme mon deuxième père dans la vie. Je remercie sincèrement Sophia de m'avoir aidée à corriger toutes les fautes de frappe françaises dans l'introduction de mon manuscrit, ce fut une grande bénédiction que je n'oublierai jamais.

Je tiens également à remercier mes deux rapporteurs Mas André et Le Penec Erwan. C'est un grand honneur pour moi d'avoir ma thèse évaluée par ces admirables experts statistique. Ma gratitude s'adresse également à Laurent-Bonneau Béatrice, Fromont-Renoir Magalie, Pham Ngoc Thanh Mai et Risser Laurent qui ont bien voulu faire partie du jury.

Je tiens à remercier le financement assuré par projet *HEARVOL* de CNRS pour la bourse de 3 ans de thèse et ainsi la LabEx de CIMI pour la bourse de Master 2 à Toulouse.

Je garde un excellent souvenir de mes trois années passées au sein de l'IMT. À ce titre, je voudrais remercier les membres et ex-membres du laboratoire et l'école doctorale pour ce qu'ils m'ont apporté. Je commence par remercier chaleureusement les membres permanents. Agnès R. pour ses explications parfois kafkaïennes du fonctionnement de l'administration universitaire; Tamara A. pour sa rédaction du

livre d'accueil des doctorants qui sera très utile aux futurs arrivants; Marie-Line D. pour la procédure du bureau et de l'équipement; Martine L. pour toute aide dans les documents administratifs de soutenance de thèse et toujours bien comprendre mes problèmes pour m'encourager à chaque rencontre avec elle; Célia V. pour toutes les procédures contractuelles et les documents relatifs aux étudiants internationaux à l'INSA; Linda C. pour la procédure d'admission.

Je remercie sincèrement les membres de l'IMT qui m'ont aidé et m'ont beaucoup motivée pendant mon doctorat: Sabine pour partager ses sentiments et la culture à travers de nombreux voyages, je n'oublierais jamais le soir de Noël 2017 où nous avons organisé une fête et distribué de la nourriture aux sans abri; Agnès pour ses petits cadeaux, qui m'ont beaucoup encouragée, ils sont vraiment génial; Xavier pour m'avoir aidé beaucoup de fois quand j'ai rencontré des difficultés dans mon travail en été 2017, il m'a encouragée mentalement, un ami qui aime beaucoup la langue vietnamienne et que je n'oublierai jamais; François, il est devenu que je suis très heureuse de connaître de notre leçon cuisine vietnamienne chez Sabine; Laurent qui m'a beaucoup aidé dans le domaine de l'analyse numérique, de l'analyse des données, de l'orientation et du partage de ses expériences de recherche et couragée des difficultés dans ma recherche; Magali, une bonne amie qui a répondu à beaucoup mes questions, et ses expériences de la vie et pour des supers souvenirs ensemble en particulier par sa fête d'anniversaire à Tarbes; Franzisca pour des souvenirs inoubliables dans la salle café pendant la pause repas et des pique-niques organisés par Sabine et Camille.

Je remercie sincèrement les membres de laboratoire de Faculté de Médecine Purpan, qui m'a beaucoup aidé dans le temps que j'ai travaillé dans ce laboratoire. José, Jean pour m'aider beaucoup dans la recherche sur les données cochléaires; Ntombi pour beaucoup de bons moments et de partages, nous réussirons grâce à nos efforts dans un pays étranger; Lubo pour des bons moments à l'heure du déjeuner.

Je remercie sincèrement les anciens doctorants de l'IMT: Pierre, calme mais enthousiaste avec d'autres; Andressa avec plein de souvenirs ensemble et de partage d'articles personnels; Antoine, toujours très drôle et gentil; Fanny, toujours enthousiaste pour aider les étudiants internationaux; Valentin, toujours calme mais enthousiaste et gentil.

Je remercie profondément les doctorants à l'IMT avec qui j'ai partagé trois ans de thèse riche en émotion: William pour tous les super moments que nous avons passé pendant plus deux ans assis au même bureau (chaque fois j'irai à la plage, je me souviendrai de toi lorsque tu dormais et réfléchissais dans le bureau :D); Camille, une meilleure amie, merci spécial à toi, de très bonne compagnie au cours des deux dernières années, c'est qui m'enseigne patiemment le français et qui est la seule à comprendre ce que je veux dire même si je parle mal le français. Je ne ferai jamais nos drôles d'actions au bureau, ma secrétaire madame Camille :D; Jose, un bon ami qui m'a donné des bons conseils et d'excellentes discussions sur des questions statistiques, et je souviens des discussions politiques avec toi aussi, elles étaient si drôles; Maylis, une fille enthousiaste, toujours prête à organiser des activités pour les doctorants, aimable et très gentille. Merci beaucoup de m'avoir

---

invité à venir dans votre (Maylis + Valentin) chambre le soir du jouer avec des doctorants et je suis très heureux quand on peut parler beaucoup de choses quand j'utilise plus le français; An, une amie, une grand soeur pour partager toutes mes difficultés ou parfois partager une chambre, et des bonnes voyages avec votre famille à Font Ramer et faire des plantes dans ton jardin, il est très étrange que nous soyons venus en France dans le même vol et que nous défendions également la thèse à la même date; Mickael, un garçon très sympa et gentil, merci beaucoup de m'avoir envoyé beaucoup d'informations utiles sur les emplois après le doctorat - et c'est qui j'ai beaucoup parlé en cours de M2R; Fabien, toujours drôle en apparence et qui a toujours laissé un coeur au tableau dans n'importe quel bureau qu'il visite; Eva, une fille de caractère, qui a répondu à beaucoup de mes questions; je me rappellerai toujours de notre soir en bateau sur la Seine, avec toi, Camille et Anouar; Paula, une fille très calme mais très gentille; Phuong, une ami, un grand frère toujours très positif avec qui j'ai partagé mes difficultés au cours des deux dernières années; Baptiste, qui est très sympa et toujours enthousiaste d'aider les doctorants; Anouar, un garçon très sympa, qui j'ai suivi le cours de statistique asymptotique et que je revois encore à Toulouse quand il travaille avec Beatrice; Laetitia, Kuntal et Clement pour de belles conversations dans les réunions de doctorants.

Je remercie aussi les doctorants à laboiraire de Faculté de Médecine Purpan avec qui j'ai partagé les derniers mois de ma thèse riche en émotion: Ntombi, une fille douce, qui j'ai partagé beaucoup beaucoup d'émotion dans le dernière partie de thesis, nous avons tenté les même cibles pour le futur parce que nous sommes des filles internationales qui fait le doctorat en France; Lubo, Aurau,... pour des bons moments à l'heure du déjeuner et j'ai eu l'occasion de rencontrer et patarger l'expérience dans le laboratoire de biologie.

Je remercie mes professeurs vietnamiens pour leur soutien constant et leurs encouragements: Nguyen Thanh Long, qui m'a toujours donné des conseils; co Phan Tu, qui m'a réconforté de nombreuses fois quand je me sentais loin de ma famille; Nguyen Tien Zung, qui m'a toujours aidée et donné des conseils utiles et qui m'a aidé beaucoup de choses dans la période difficile de ma thèse. Il est un professeur extraordinaire qui donne toujours ce qu'il y a de mieux aux étudiants vietnamiens qui font des mathématiques à Toulouse

Je remercie mes amis français que j'ai connus en France: Benjamin pour son aide enthousiaste dans la recherche d'un emploi et ses conseils sur la vie en France. Grégory avec beaucoup de bons souvenirs quand j'ai cherché un colocataire pour trouver une motivation de départ pour apprendre le français. Il m'a beaucoup aidé patiemment à corriger chaque mot français lorsque j'ai essayé de lui raconter une histoire que je lisais tous les jours. Merci beaucoup à toi, un bon ami de tous les temps. Laure, une bonne amie qui j'ai connu quand j'ai habite dans une chambre colocataire apres avec Gregory. :) Tania pour tous m'aide dans le voyage de retour de Toulouse au Vietnam, où j'ai perdu ma poche avec de l'argent et des documents.

Merci aux relatifs au Vietnam qui m'ont toujours encouragé au cours des quatre dernières années en France: Ba Ngoai, Ba Noi, ba Can, ba Duyen, 2 bac Cuong-Huong, 2 bac Vien, co chu Hien-Luc, anh chi Tai-Ly, chi Lan, chi Hoa, anh Lam,

anh Chinh, chi Loc... va tat ca ho hang hai ben noi ngoai.

Je tiens à remercier mes amis qui vivent au Vietnam et différentes régions mais qui me donnent toujours des encouragements pendant quatre ans: co Thanh, anh Ha, chi Nhung, anh Minh, anh Khanh, anh Thanh, em Y, Minh Trung, Minh Duy, Nhat Truong, Bich Thuan, Thanh Tien, Thanh Viet, Quan On, Cam Tu, Phan Thuy... et em Phan Thu; amis du lycée: Thanh Huong, Nhung, My Luong, Duc Thai, Tien Hung, Huy Quy, Thanh Phuong, Lai Phuong, Hoang Nhi...; amies j'ai connu en France: anh Thieu, em Phan Anh, em Thuy Orleon,... Ils me donnent toujours des encouragements par le biais de messages même si nous restons dans un fuseau horaire différent. Je tiens à remercier tout particulièrement Thanh Nhan, Phuoc Nhat et Tien Dat, qui ont passé du temps à venir à Toulouse pour ma thèse et des bons souvenirs de nos beaux voyages en France. Ensuite, je tiens à remercier sincèrement anh Song Phuong, qui se souvient toujours de tous mes événements importants et m'encourage beaucoup par le biais de courriels, de messagers.

Pendant quatre ans à Toulouse, j'ai été très heureux de reconstruire et de faire connaissance avec beaucoup des bonnes amies qui m'ont partagé de nombreux souvenirs agréables: anh chi Hoang-Thom avec les souvenirs inoubliables et anh Hoang pour m'a beaucoup aidé les connaissances statistiques pendant dans les deux premières années à Toulouse, chi Mong Ngoc pour m'a beaucoup aidé à connaître plus d'amis français à Toulouse, et partagé les difficultés en première année de ma thèse; anh chi Hang-Dat, je suis reconnaissant pour toutes vos aides pendant quatre ans en restant à Toulouse, en particulier, à la fin de la première année de doctorat pour m'aider en passant bien les défis dans les travaux et de l'esprit. Merci beaucoup pour votre soutien, vos conseils et vos organisations formidables à vous joindre à de belles fêtes avec un groupe vietnamien à Toulouse. Ce sont mes souvenirs inoubliables dans cette ville; chi Nga, une bonne amie, une grande sœur pour partager toutes des choses dans ma vie pendant quatre ans, qui toujours sympathie avec moi dans toutes les situations. Nous sommes des filles fortes et j'espère qu'un jour je te verrai heureuse pour toujours dans tous tes choix; anh chi Chien-Khanh, pour supers memories dans le voyage au Andorra des moments agréables avec Joss également, des encouragements et le partage de nombreux cadeaux du Vietnam; anh chi Khoa-An, pour les supers souvenirs dans le voyage visiter Font Romeu et les souvenirs inoubliables dans des périodes difficiles pendant quatre ans à Toulouse; anh chi Thuyen-Han pour tout les incroyables memories ensemble dans la dernière année de thèse, merci anh chi pour beaucoup soutien, conseils dans ma vie, c'est dommage que anh chi n'ira pas dans ma soutenance de thèse; anh chi Hung-Nguyen, pour les encouragements dans ma thèse; gia dinh Hau-anh Chuc et Thuy, pour les encouragements quand j'ai des difficultés et beaucoup des bons souvenirs; anh chi Phong-Ngoc, anh chi Chinh-Huyen pour les bons souvenirs avec un groupe Vietnamien faisant des maths à Toulouse; Merci beaucoup pour beaucoup de bons échanges avec vous deux et donnez-moi toujours des encouragements; Trang-Tu: nous reverrons des cartes postales échangées à chaque voyage les deux dernières années et penserons les uns aux autres; Hoang Phuong, pour les bons souvenirs dans les deux premières années à Toulouse; Hieu, pour beaucoup de bons souvenirs à travers de belles conversations de

trois ans; em Tin, qui m'a encouragé toujours par des messages chaleureux; Thanh Trung, Dinh Lan, Hoang Anh, pour les bons souvenirs avec un groupe Vietnamien à Toulouse.

Je immense merci ensuite à mes deux amies de toujours. Je veux bien sûr parler de Ngoc (rencontrée en 6 ans déjà), Nous avons passé ensemble de nombreux événements tristes et heureux depuis le jour où j'étais à HoChiMinh ville. Aucun mot ne peut présenter mon bonheur quand j'ai toujours une amie formidable à mes côtés dans toutes les situations de ma vie. C'est vraiment triste quand vous ne pouvez pas être présent dans ma soutenance de thèse, mais vous êtes toujours dans mon esprit comme un meilleur ami de tous les temps <3. Josselin, qui est une position spéciale dans mon coeur. Merci Josselin pour beaucoup de voyages inoubliables ensemble, nous avons fait beaucoup de bons souvenirs ensemble et pour tes conseils comme un grand frère dans ma vie en France, qui m'a appris à profiter au maximum de chaque instant de la vie. Je n'oublierai jamais les moments de montagne où nous faisons du trekking, premières expériences de ma vie à travers nos beaux voyages <3.

Enfin, je remercie du fond du coeur mes parents et mon frère au Vietnam, qui sont toujours reste près de moi et m'encourager sans conditions. C'était vraiment difficile parfois de rester loin de la famille, mais ils sont toujours proches de moi dans mon coeur pour toujours. Il n'y a pas de mots pour présenter ma grande émotion à ma famille. C'est un sentiment sacré qui ne peut être exprimé avec des mot, sje vous aime tous... Et je ne sais pas où je vais rester pour me reposer, mais le Vietnam est toujours une partie importante de ma vie.

# Chapitre 1

## Introduction générale

### Contents

---

<b>1.1 Motivation</b> . . . . .	<b>11</b>
1.1.1 Le modèle de régression . . . . .	12
1.1.2 Test d'hypothèses en régression non paramétrique . . . . .	13
<b>1.2 Objectifs</b> . . . . .	<b>14</b>
<b>1.3 Présentation des résultats</b> . . . . .	<b>16</b>
1.3.1 Régression dans un RKHS pour des distributions: application à la prévision de l'âge en fonction des courbes d'oto-émission . . . . .	16
1.3.2 Tests sur courbes . . . . .	20

---

### 1.1 Motivation

Dans la modélisation statistique, l'analyse de régression est un ensemble de processus statistiques permettant d'estimer les relations entre les variables. Il comprend de nombreuses techniques pour modéliser et analyser plusieurs variables lorsque l'accent est mis sur la relation entre une variable dépendante et une ou plusieurs variables indépendantes (ou "prédicteurs"). Plus précisément, l'analyse de régression aide à comprendre comment la valeur typique de la variable dépendante (ou "variable critère") change lorsque l'une des variables indépendantes est modifiée, tandis que les autres variables indépendantes sont maintenues fixes. Voir par exemple [Azaïs \(2006\)](#); [Kutner, Nachtsheim, and Neter \(2004\)](#) et les références qui y figurent. Le plus souvent, l'analyse de régression estime l'espérance conditionnelle de la variable dépendante en fonction des variables indépendantes, c'est-à-dire la valeur moyenne de la variable dépendante lorsque les variables indépendantes sont fixes. Moins communément, l'accent est mis sur un quantile ou un autre paramètre d'emplacement de la distribution conditionnelle de la variable dépendante étant donné les variables indépendantes. Dans tous les cas, une fonction des variables indépendantes



appelée fonction de régression doit être estimée. L'analyse de régression est largement utilisée pour la prévision, son utilisation englobant largement le domaine de l'apprentissage automatique. L'analyse de régression est également utilisée pour comprendre quelles variables indépendantes sont liées à la variable dépendante et pour explorer les formes de ces relations. De nombreuses techniques d'analyse de régression ont été développées. Les méthodes familières telles que la régression linéaire et la régression des moindres carrés ordinaires sont paramétriques, en ce sens que la fonction de régression est définie en termes d'un nombre fini de paramètres inconnus estimés à partir des données. La régression non paramétrique fait référence à des techniques permettant à la fonction de régression de se situer dans un ensemble spécifique de fonctions, qui peut être de dimension infinie.

### 1.1.1 Le modèle de régression

Le modèle de régression le plus général est  $y = f(x) + \sigma\epsilon$ , qui indique simplement que la valeur de  $y$  (les réponses) dépend de la valeur de  $x$  (les prédicteurs). On peut supposer que le terme d'erreur aléatoire  $\epsilon$  est normalement (et indépendamment) distribué, de moyenne nulle et de variance  $\sigma^2$ ;  $f$  est la fonction de régression qui doit être estimée. Dans divers domaines d'application, différentes terminologies sont utilisées à la place des variables dépendantes et indépendantes.

Un large éventail de modèles de régression ont été proposés lorsque  $x$  et/ou  $y$  prennent la forme de courbes ou de fonctions. Il est devenu classique de distinguer trois situations de base dans ce domaine de régression fonctionnelle (Ramsay (2005), Chiou, Müller, and Wang (2004)): le premier, les réponses et les prédicteurs sont des fonctions; le deuxième, les réponses sont des scalaires et les prédicteurs sont des fonctions; et enfin les réponses sont des fonctions et les prédicteurs sont des scalaires. Dans ce cadre, différentes stratégies ont été étudiées et utilisées dans plusieurs domaines d'application concernant l'analyse de données fonctionnelles (Neter, Kutner, Nachtsheim, and Wasserman (1996), Ramsay and Silverman (2007)). Le cadre RKHS (Reproducing Kernel Hilbert Space) est récemment devenu populaire pour étendre les résultats de la théorie de l'apprentissage statistique dans le contexte de la régression de données fonctionnelles, ainsi que pour développer des procédures d'estimation de prédicteur fonctionnel  $f$  comme dans Preda (2007) et Kadri, Duflos, Preux, Canu, and Davy (2010) par exemple. Ce cadre est particulièrement important dans le domaine de la théorie statistique de l'apprentissage en raison du théorème du Représentant, qui stipule que chaque fonction peut être écrite comme une combinaison linéaire de la fonction noyau évaluée aux points d'apprentissage (Berlinet and Thomas-Agnan (2011)). On obtient ainsi un moyen d'estimer des fonctions sur un large éventail d'espaces fonctionnels.

Les distributions de probabilités constituent un autre domaine d'intérêt pour la régression et il est devenu une préoccupation majeure ces dernières années. Semblable à la régression fonctionnelle, le modèle de régression de distribution peut être classé en trois types selon que les réponses ou les prédicteurs sont des distributions ou des scalaires. Dans notre cadre, nous cherchons à résoudre le problème de ré-

gression avec des entrées appartenant à des espaces de distribution de probabilité dont les sorties sont des valeurs réelles. Une approche commune de ce problème consiste à examiner le plongement dans un espace de Hilbert à l'aide de noyaux, en utilisant par exemple le plongement moyen comme dans [Smola, Gretton, Song, and Schölkopf \(2007\)](#). Voir [Muandet, Fukumizu, Dinuzzo, and Schölkopf \(2012\)](#) pour d'autres références. Les études de régression avec entrées de type distribution et réponses réelles sont l'un des sujets de recherche les plus populaires dans ce domaine, par exemple [Póczos, Rinaldo, Singh, and Wasserman \(2013\)](#) et [Oliva, Neiswanger, Póczos, Schneider, and Xing \(2014\)](#).

### 1.1.2 Test d'hypothèses en régression non paramétrique

Une hypothèse statistique, parfois appelée analyse de données de confirmation, est une hypothèse qui peut être testée sur la base de l'observation d'un processus modélisé via un ensemble de variables aléatoires. Un test d'hypothèse statistique est une méthode d'inférence statistique. Généralement, on compare deux ensembles de données statistiques ou on compare un ensemble de données obtenu par échantillonnage à un ensemble de données synthétiques provenant d'un modèle idéalisé. Une hypothèse est proposée pour la relation statistique entre les deux ensembles de données. Cette hypothèse est comparée en tant qu'alternative à une hypothèse nulle idéalisée ne proposant aucune relation entre deux ensembles de données. La comparaison est considérée comme statistiquement significative si la relation entre les ensembles de données constitue une réalisation improbable de l'hypothèse nulle selon une probabilité de seuil, le niveau de signification. Les tests d'hypothèses sont utilisés pour déterminer quels résultats d'une étude conduiraient à un rejet de l'hypothèse nulle pour un niveau de signification prédéfini. Le processus de distinction entre l'hypothèse nulle et l'hypothèse alternative est facilité en considérant deux types d'erreurs conceptuels. Le premier type d'erreur se produit lorsque l'hypothèse nulle est rejetée à tort. Le second type d'erreur se produit lorsque l'hypothèse nulle n'est pas rejetée à tort.

Dans cette thèse, nous nous sommes intéressés aux tests d'hypothèses pour des modèles de régression non paramétriques. Ce thème a été considéré dans plusieurs articles [King \(1988\)](#), [Hardle and Marron \(1990\)](#), [Hall and Hart \(1990\)](#), [King, Hart, and Wehrly \(1991\)](#) et [Delgado \(1992\)](#). Les tests d'absence d'effet dans la régression non paramétrique sont étudiés dans [Eubank and LaRiccia \(1993\)](#). Dans le document de [Spokoiny et al. \(1996\)](#), les auteurs considèrent le cas particulier où  $\sigma$  est supposé connu. Ils proposent des tests permettant d'atteindre les taux minimax de tests pour un large éventail de classes de Besov. [Baraud, Huet, Laurent, et al. \(2003\)](#) proposent un test, basé sur des méthodes de sélection de modèle, pour tester si l'espérance d'un vecteur gaussien à  $n$  composantes indépendantes appartient à un sous-espace linéaire de  $\mathbb{R}^n$  par rapport à une alternative non paramétrique. La procédure de test est disponible lorsque la variance des observations est inconnue et ne dépend d'aucune information préalable sur l'alternative. Ils obtiennent des taux de test optimaux allant jusqu'à un facteur  $\log n$  possible sur différentes classes d'alternatives

simultanément. Plus récemment, dans un cadre de processus de Poisson, [Fromont, Lerasle, Reynaud-Bouret, et al. \(2012\)](#) et [Fromont, Laurent, Reynaud-Bouret, et al. \(2013\)](#) ont envisagé deux processus de Poisson indépendants. Ils abordent la question du test de l'égalité de leurs intensités respectives. Ils introduisent des tests basés sur une seule fonction du noyau et agrègent plusieurs tests basés sur le noyau pour obtenir des procédures de test adaptatives minimales par rapport à des alternatives basées sur les boules de Besov ou de Sobolev. Notre deuxième travail est fortement inspiré du papier de [Fromont et al. \(2013\)](#).

## 1.2 Objectifs

Comme nous l'avons évoqué, cette thèse concerne la régression du noyau à travers deux approches. L'une d'entre elles concerne l'introduction d'une nouvelle prédiction de la fonction de régression pour les modèles de régression avec des entrées de distribution. La seconde aborde la question de tester la nullité de la fonction de régression sur un modèle de régression. Nous avons appliqué notre première approche à des données biologiques. Nous avons choisi des profils auditifs (mesurés en Otoémissions acoustiques évoquées transitoires - TEOAEs) car ils peuvent contenir en substance des informations cachées sur des traits de vie individuels (classes d'âge) et constituent le candidat idéal pour les analyses de régression. Plus précisément, nous détaillons ci-dessous les deux objectifs de cette thèse.

Pour le premier objectif de cette thèse, nous considérons les noyaux construits en utilisant la distance de Wasserstein (voir [Villani \(2008\)](#)). Dans le papier de [Kolouri, Zou, and Rohde \(2015\)](#), les auteurs construisent un noyau de Wasserstein en tranches basé sur la distance de Wasserstein, qui s'est révélé être un noyau défini positif pour des distributions absolument continues. Plus généralement, pour toutes les distributions, les auteurs de [Bachoc, Gamboa, Loubes, and Venet \(2017\)](#) ont construit une famille de noyaux définis positifs, définis avec la distance de Wasserstein pour les distributions sur la ligne réelle. De plus, un noyau étroitement lié à ceux-ci a été précédemment étudié via les "noyaux universels" dans [Steinwart \(2001\)](#) afin d'atteindre la richesse de RKHS. Dans le cadre de [Christmann and Steinwart \(2010\)](#), les auteurs prouvent l'universalité du noyau RBF de type gaussien. De plus, combinant le noyau défini positif construit par [Bachoc et al. \(2017\)](#) avec le noyau Gaussien radial dans [Christmann and Steinwart \(2010\)](#), nous considérons un noyau universel basé sur la distance de Wasserstein entre deux distributions de probabilité sur un espace compact. Plus précisément, cela correspond aux deux étapes prévues dans la suite de notre travail, [Bui, Loubes, et al. \(2018\)](#). Premièrement, nous considérons l'espace des distributions comme une variété munie des distances de Wasserstein. Deuxièmement, nous avons abordé la question des régressions sur les espaces fonctionnels RKHS, et l'avons appliqué à l'étude des liens entre les courbes de sensibilité auditive et les classes d'âge.

Le contexte applicatif de notre modèle concerne l'exploitation des données auditives, par le biais des otoémissions acoustiques. Les otoémissions sont des fréquences

émises par l'oreille sur une gamme de fréquences suite à sa stimulation par de multiples (ici 256) fréquences. Ces stimuli vont faire vibrer les cellules ciliées externes, situées le long de la membrane basilaire cochléaire, qui vont produire un son en réponse. Ce phénomène, prédit par [Kemp \(1978\)](#), a été démontré expérimentalement par le physicien David Kemp grâce au perfectionnement de micros ultrasensibles. Chez l'homme, les TEOAEs ont été très utilisées pour évaluer la sensibilité auditive, mais des études fines portant sur un lien potentiel entre ces courbes de sensibilités et d'autres paramètres biologiques n'ont pas été menées jusque là. C'est en ce sens que l'objet d'application de notre modèle est original. Les premières études sur les TEOAE ont été présentées par [Johnsen, Bagi, and Elberling \(1983\)](#); [Kemp and Ryan \(1991\)](#). La différence dans la structure des courbes TEOAE varie avec l'âge. Par exemple, les amplitudes de TEOAE chez les nourrissons peuvent être plus grandes que chez l'adulte, dans des conditions d'enregistrement comparables. Les composantes principales de fréquences peuvent être plus grandes également [Prieve, Fitzgerald, and Schulte \(1997b\)](#). Le niveau de réponse total au TEOAE diminue avec l'âge après les 2 premiers mois de la vie, [Liu and Wang \(2012\)](#). Pour cela, nous avons considéré les données TEOAEs dans un cadre utilisant la géométrie des données. Les TEOAE ont été considérées comme des objets fonctionnels ayant une densité indiquant à quelle fréquence une gamme de fréquences a été activée. La recherche de la relation entre la sensibilité auditive et les groupes d'âge est traitée en appliquant ce cadre, nous avons prédit l'âge des TEOAE dans la population humaine mais également en utilisant des données non-humaines, [Fowler, Chiasson, Leslie, Thomas, Beasley, Kemnitz, and Weindruch \(2010\)](#).

Pour le second objectif de cette thèse, nous avons considéré un modèle de régression avec des entrées de type fonction et avons abordé la question du test de la nullité de la fonction de régression, [Bui \(2019\)](#). Nous proposons tout d'abord une nouvelle procédure de test unique basée sur un noyau symétrique général et une estimation de la variance des observations. Au lieu de considérer un seul noyau particulier, nous considérons une collection de noyaux et la collection correspondante de tests, chacun avec un niveau de signification adapté. Nous rejetons ensuite l'hypothèse nulle lorsqu'il existe au moins un des tests de la collection qui rejette l'hypothèse nulle. Les procédures de test agrégées sont construites pour être de niveau  $\alpha$  et l'erreur de seconde espèce due à l'agrégation, lorsque elle est inévitable, est aussi petite que possible. Les propriétés des procédures sont examinées par simulation numériques. Les résultats sur données synthétiques générées aléatoirement montrent que les tests proposés maintiennent les taux d'erreur souhaités sous l'hypothèse nulle et ont une bonne puissance sous l'alternative. Dans l'état actuel de mes travaux, il n'est pas possible d'appliquer les tests d'hypothèse proposés ici aux otoémissions parce que les données ne contiennent pas nécessairement les points (1.3.7) qui permettent d'estimer la variance des observations.

## 1.3 Présentation des résultats

Cette section est composée de deux parties indépendantes décrites dans des articles soumis ou en cours de finalisation. Nous avons brièvement présenté ici nos résultats de chaque partie.

### 1.3.1 Régression dans un RKHS pour des distributions: application à la prévision de l'âge en fonction des courbes d'oto-émission

Dans cette partie, nous étudions des modèles de régression avec des entrées de distribution. Ce travail réalisé avec Loubes, J.M., Risser, L. et Balaesque, P. a fait l'objet de l'article [Bui et al. \(2018\)](#).

Nous voulions résoudre le problème de régression avec des entrées appartenant à des espaces de distribution de probabilités et des sorties réelles. En particulier, nous considérons le modèle

$$y_i = f(\mu_i) + \sigma \epsilon_i, \quad (1.3.1)$$

où  $\{\mu_i\}_{i=1}^n$  sont les distributions de probabilité sur  $\mathbb{R}$ ,  $\{y_i\}_{i=1}^n$  sont des nombres réels,  $\epsilon_i$  représente un bruit gaussien indépendant et identiquement distribué et  $\sigma$  est le niveau de bruit. Comme dans les modèles de régression classiques, nous avons besoin d'observations  $\{(\mu_i, y_i)\}_{i=1}^n$ .

Pour trouver une estimation de la fonction de régression inconnue  $f$ , nous voulions utiliser le cadre de la minimisation du risque empirique régularisé sur l'espace de Hilbert à noyau reproduisant (RKHS). Plus particulièrement, nous cherchons à trouver la solution du problème de minimisation suivant

$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \left( \sum_{i=1}^n |y_i - f(\mu_i)|^2 + \lambda \|f\|_{\mathcal{F}}^2 \right), \quad (1.3.2)$$

où  $\lambda \in \mathbb{R}^+$  est le paramètre de régularisation et  $\mathcal{F}$  est un espace de Hilbert à noyau reproduisant.

Comme nous l'avons présenté dans la motivation de cet article, nous considérons le noyau Gaussien radial présenté dans [Christmann and Steinwart \(2010\)](#). Il est défini comme suit

**Proposition 1.1.** *Soit  $X$  un espace métrique compact et  $\mathcal{H}$  un espace de Hilbert séparable tel qu'il existe une application continue et injective  $\rho : X \rightarrow \mathcal{H}$ . Pour  $\gamma > 0$ , le noyau Gaussien radial  $k_\gamma : X \times X \rightarrow \mathbb{R}$  est le noyau universel, défini pour tout  $x, x' \in X$  par*

$$k_\sigma(x, x') := \exp(-\sigma^2 \|\rho(x) - \rho(x')\|_{\mathcal{H}}^2).$$

Rappelons ici le résultat suivant, issu de l'article de [Bachoc et al. \(2017\)](#).

**Théorème 1.1.** Soit  $k_\Theta : \mathcal{W}_2(\Omega) \times \mathcal{W}_2(\Omega) \rightarrow \mathbb{R}$  avec le paramètre  $\Theta := (\gamma, H, l)$  tel que  $\gamma \neq 0$  et  $l > 0$  défini comme

$$k_\Theta(\mu, \nu) := \gamma^2 \exp\left(-\frac{W_2^{2H}(\mu, \nu)}{l}\right). \quad (1.3.3)$$

Alors pour  $0 < H \leq 1$ ,  $k_\Theta$  est un noyau défini positif.

Ces deux résultats permettent de construire un noyau universel de type Gaussien basé sur la distance de Wasserstein:

**Théorème 1.2.** Soit  $k_\Theta : \mathcal{W}_2(\Omega) \times \mathcal{W}_2(\Omega) \rightarrow \mathbb{R}$  avec le paramètre  $\Theta := (\gamma, H, l)$  tel que  $\gamma \neq 0$ ,  $H = 1$  et  $l > 0$  défini comme

$$k_\Theta(\mu, \nu) := \gamma^2 \exp\left(-\frac{W_2^{2H}(\mu, \nu)}{l}\right). \quad (1.3.4)$$

Alors le noyau  $k_\Theta$  est universel.

Ensuite, nous pourrions construire facilement un RKHS  $\mathcal{F}$  à partir d'un noyau universel. Enfin, en utilisant le cadre RKHS par minimisation du risque empirique régularisé (1.3.2) impliquant des distributions dans les RKHS (Muandet et al. (2012)), nous avons résolu le problème de la fonction inconnue  $f$  dans le modèle de régression avec entrée de distribution. La fonction de régression prédite a été exprimée par

$$\hat{f} : \mu \mapsto \hat{f}(\mu) := \sum_{j=1}^n \hat{\alpha}_j k_\Theta(\mu, \mu_j), \quad (1.3.5)$$

où  $\hat{\alpha} = (C + \lambda I)^{-1} Y$ , avec  $\alpha = (\alpha_1, \dots, \alpha_n)^T$ ,  $Y = (y_1, \dots, y_n)^T$  et  $C = \{(c_{ij})\}_{i,j=1}^n$ ,

$$c_{ij} = \gamma^2 \exp\left(-\frac{W_2^2(\mu_i, \mu_j)}{l}\right).$$

Nous avons utilisé l'erreur quadratique moyenne (RMSE) comme critère pour montrer la différence de qualité entre notre modèle et d'autres modèles. Dans les entrées de type distribution en dimension 1, nous avons comparé notre modèle avec deux autres modèles. L'un d'eux noté par "Legendre" correspond aux modèles de régression avec entrée de densité lisse et l'autre noté par "Histogram" est introduit pour l'entrée d'histogramme approximatif. Le critère de qualité est que RMSE devrait être minimale. Donc, notre modèle surpasse clairement les autres modèles. On constate dans le Table 1.1 que.

En outre, nous avons comparé la qualité de notre noyau d'une sorte de noyau très proche du notre présenté dans Kolouri et al. (2015); Kolouri, Rohde, and Hoffmann (2018). Nous avons noté notre modèle en dimension 2 par "Wasserstein2D" et l'autre par "SlicedWasserstein2D". Dans ce cas, nous avons fait référence à un noyau universel basé sur la distance de Wasserstein entre deux distributions Gaussiennes en dimension 2 décrit dans Bachoc, Suvorikova, Loubes, and Spokoiny (2018). Dans la Table 1.2, en se basant sur les valeurs RMSE minimales, le modèle "Wasserstein2D" surpasse nettement le modèle "SlicedWasserstein2D" lorsqu'il existe beaucoup de bruit dans le modèle de régression.

model	RMSE
"Wasserstein"	<b>0.04</b>
"Legendre" order 5	<b>0.15</b>
"Legendre" order 10	<b>0.11</b>
"Histogram"	<b>0.12</b>

Table 1.1: Les valeurs RMSE pour les modèles de régression de distribution "Wasserstein", "Legendre" et "Histogram".

$\sigma$	"Wasserstein2D"	"SlicedWasserstein2D"
1	<b>0.06</b>	<b>0.06</b>
5	<b>0.09</b>	<b>0.33</b>
10	<b>0.1</b>	<b>0.91</b>

Table 1.2: Les valeurs RMSE pour les modèles de régression de distribution "Wasserstein2D" and "Sliced-Wasserstein2D" en fonction du niveau de bruit  $\sigma = 1, 5, 10$ .

### Application à l'évolution de la sensibilité auditive

Nous avons examiné 48 profils TEOAE correspondant à 48 individus (âgés de 15 à 50 ans), dans la Figure 1.1. Chaque observation est associée à une courbe d'oto-émission qui traduit la réponse de la cochlée à plusieurs stimuli fréquences (256 fréquences). Le niveau de réponse dépend de chaque individu et chaque stimulus doit être normalisé, mais la réaction de chaque individu est caractéristique de ses caractéristiques physiologiques. Ainsi, à chaque individu est associée une courbe qui, après normalisation, est considérée comme une distribution  $\mu$  décrivant la répartition des réponses pour différentes fréquences allant théoriquement de 0 à 10 kHz.

En utilisant le modèle de régression de distribution, nous avons étudié la relation entre les réponses évoquées de clics couvrant la plage de fréquences allant jusqu'à 10 kHz et l'âge. Nous avons très bien prédit l'âge de chaque individu en appliquant notre fonction de régression (Figure 1.2). En particulier, on obtient une bonne prédiction sur l'échantillon appartenant à la classe d'âge [15, 35].

### Perspective

Dans ce travail, nous avons étudié avec succès la courbe TEOAE de chaque individu dans la population humaine sous forme de distribution après normalisation. Nous avons ensuite étudié la relation entre l'âge et la courbe TEOAE. L'effet de l'âge sur l'audition est profondément lié au changement de cochléaire. Il s'agit d'une



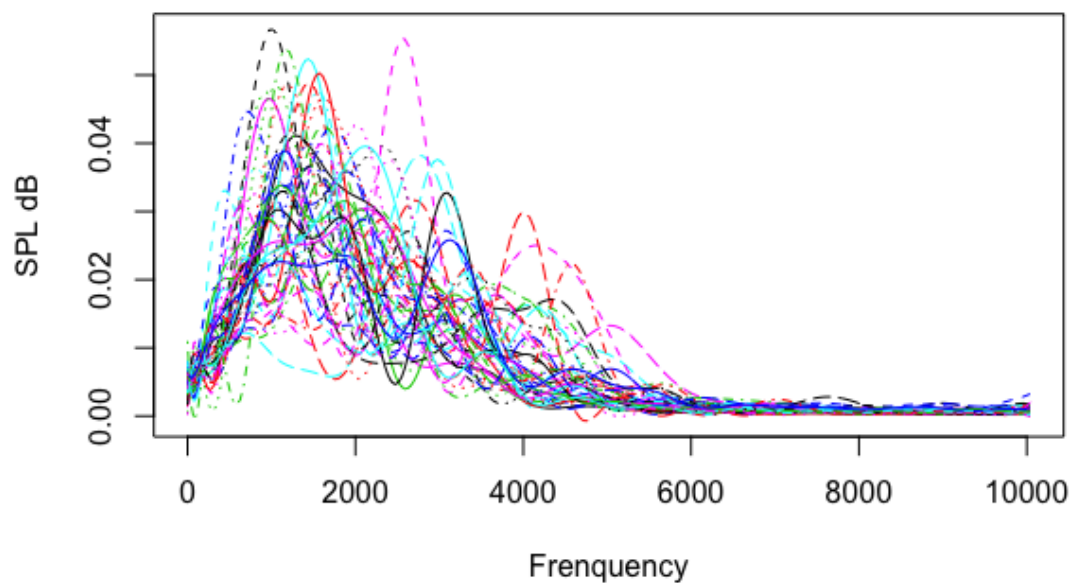


Figure 1.1: Courbes d'oto-émission. 48 courbes de TEOAE sur la gamme de fréquence théorique maximale (0Hz-10kHz).

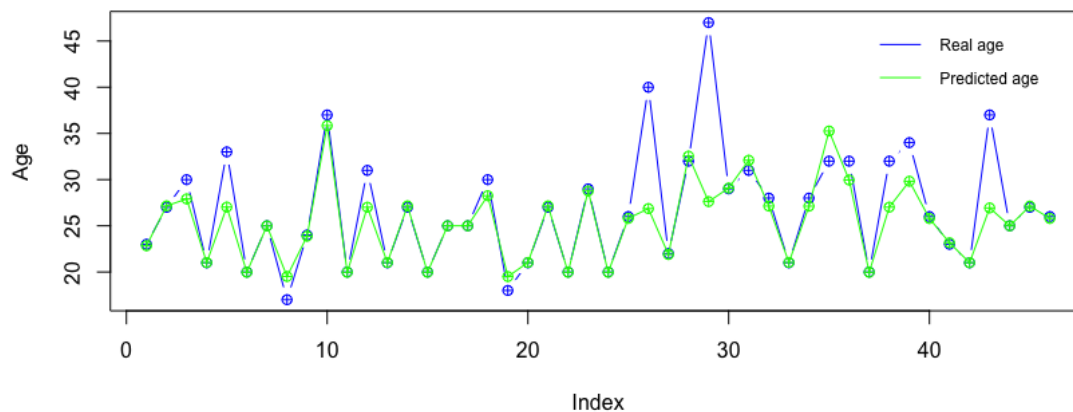


Figure 1.2: L'âge exact et l'âge prédit.

nouvelle approche intéressante dans le domaine du traitement statistique de cette donnée biologique. Nous pensons que notre article aborde un problème important pour les experts en science des données souhaitant prédire les problèmes de régression avec les distributions de probabilité en entrée. Nous souhaitons en outre développer d'avantage l'application de ce modèle à d'autres espèces pour lesquelles les profils sont bien plus hétérogènes afin de tester la transférabilité de la méthode à d'autres



sujets. L'extension de ce travail à des distributions en dimension supérieure devrait être abordée dans un travail ultérieur, en utilisant par exemple comme noyau celui qui se trouve dans [Bachoc et al. \(2018\)](#).

### 1.3.2 Tests sur courbes

Cette partie concerne les tests d'hypothèse. Elle a fait l'objet de l'article [Bui \(2019\)](#).

#### Problème et test statistique proposés

Considérons  $(X_i, Y_i)_{1 \leq i \leq n}$  qui obéissent au modèle de régression décrit comme suit,

$$Y_i = f(X_i) + \sigma \epsilon_i, \quad i = 1, \dots, n. \quad (1.3.6)$$

Nous supposons que  $X = (X_1, X_2, \dots, X_n)$  sont des variables aléatoires réelles i.i.d à valeurs dans un ensemble mesurable  $E$  telle que  $[0, 1] \subset E \subset \mathbb{R}$  de densité  $\nu$  par rapport à la mesure de Lebesgue sur  $E$  et  $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$  sont des variables aléatoires gaussiennes standard, i.i.d indépendantes de  $(X_1, X_2, \dots, X_n)$ . La fonction  $f$  est supposée être dans  $\mathbb{L}^2(E, d\nu)$ . Nous présumons également que  $\|f\|_\infty = \sup_{x \in E} |f(x)| < +\infty$ . Afin d'évaluer  $\sigma^2$ , nous avons supposé que nous avons aussi observé  $(Y'_1, \dots, Y'_n)$  qui obéissent au modèle

$$Y'_i = f\left(\frac{i}{n}\right) + \sigma \epsilon'_i, \quad i = 1, \dots, n, \quad (1.3.7)$$

où  $\epsilon' = (\epsilon'_1, \dots, \epsilon'_n)$  est indépendant de  $(X_1, \dots, X_n, \epsilon_1, \dots, \epsilon_n)$ .

Étant donné l'observation de  $(X_i, Y_i)_{1 \leq i \leq n}$ ,  $(Y'_i)_{1 \leq i \leq n}$ , nous voulons vérifier l'hypothèse nulle

$$(H_0) : f = 0,$$

contre l'hypothèse alternative

$$(H_1) : f \neq 0.$$

Dans ce travail, pour tester des hypothèses  $(H_0)$  contre  $(H_1)$  dans un modèle de régression, nous avons construit des procédures de test agrégées basées sur le noyau. Nos statistiques de test étaient basées sur une seule fonction du noyau qui pourrait être choisie soit comme un noyau de projection ou un noyau Gaussien. Nous avons proposé une estimation pour la variance inconnue  $\sigma^2$ . Nos tests étaient exactement (et pas seulement asymptotiquement) de niveau  $\alpha$ . Nous avons obtenu les conditions optimales non-asymptotiques dans l'hypothèse alternative qui garantissait que la probabilité d'erreur de type II était au plus égale à un niveau prescrit  $\beta$ . Cependant, nos tests avaient aussi pour but de régler la question du choix du noyau et des paramètres de ce noyau. Nous avons utilisé une approche d'agrégation, qui était bien connue dans les tests adaptatifs ([Baraud et al. \(2003\)](#) et [Fromont et al. \(2013\)](#)). Ce travail est fortement inspiré du papier de [Fromont et al. \(2013\)](#).

Soit  $K : E \times E \rightarrow \mathbb{R}$  une fonction du noyau symétrique: satisfaisant  $\int_{E^2} K^2(x, y) f(x) f(y) d\nu(x) d\nu(y) < +\infty$ .

Nous avons introduit le test d'hypothèse  $V_K$  défini ci-après

$$V_K = \frac{T_K}{\hat{\sigma}_n^2}, \quad (1.3.8)$$

où

$$T_K = \frac{1}{n(n-1)} \sum_{i \neq j=1}^n K(X_i, X_j) Y_i Y_j \quad (1.3.9)$$

et

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n/2} (Y'_{2i-1} - Y'_{2i})^2. \quad (1.3.10)$$

L'espérance de  $T_K$  est égale à

$$\mathbb{E}(T_K) = \langle K[f], f \rangle, \quad (1.3.11)$$

où

$$K[f](x) = \int_E K(x, y) f(y) d\nu(y), \quad \text{pour tout } x \in E,$$

et pour tout  $f, g \in \mathbb{L}^2(E, d\nu)$

$$\langle f, g \rangle = \int_E f(x) g(x) d\nu(x) \text{ et } \|f\|^2 = \langle f, f \rangle.$$

D'autre part,  $\mathbb{E}[\hat{\sigma}_n^2] = a^2 + \sigma^2$ , avec  $a^2 := \frac{1}{n} \sum_{i=1}^{n/2} \left[ f\left(\frac{2i-1}{n}\right) - f\left(\frac{2i}{n}\right) \right]^2$ .

Ainsi,  $\hat{\sigma}_n^2$  est un estimateur biaisé de  $\sigma^2$  avec pour biais  $a^2$ . Si  $f$  est une fonction régulière, ce biais est faible.

Nous traitons ci-après deux exemples particuliers de fonctions de noyau pour expliquer pourquoi nous avons choisi le test d'hypothèse  $V_K$ .

*Exemple 1.* Quand  $E = [0, 1]$ , notre premier choix pour  $K$  était une fonction noyau symétrique basée sur une famille  $\{\phi_\lambda, \lambda \in \Lambda\}$  orthonormale par rapport au produit scalaire  $\langle \cdot, \cdot \rangle$ ,

$$K(x, y) = \sum_{\lambda \in \Lambda} \phi_\lambda(x) \phi_\lambda(y). \quad (1.3.12)$$

Pour tous  $f$  dans  $\mathbb{L}^2([0, 1], d\nu)$  nous avons  $\mathbb{E}(T_K) = \langle \Pi_S(f), f \rangle$ , où  $S$  était le sous-espace de  $\mathbb{L}^2([0, 1], d\nu)$  engendré par les fonctions  $\{\phi_\lambda, \lambda \in \Lambda\}$  et  $\Pi_S$  indique la projection orthogonale sur  $S$  pour  $\langle \cdot, \cdot \rangle$ .

Ainsi, lorsque  $\{\phi_\lambda, \lambda \in \Lambda\}$  est bien choisie,  $T_K$  peut être considéré comme un estimateur pertinent de  $\|f\|^2$ .

*Exemple 2.* Quand  $E = \mathbb{R}$  et  $\nu$  est la fonction de densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ , notre deuxième choix  $K$  est un noyau Gaussien défini par

$$K(x, y) = \frac{1}{h} k\left(\frac{x-y}{h}\right), \text{ pour tous } (x, y) \in \mathbb{R}^2, \quad (1.3.13)$$

où  $k(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$ , pour tous  $u \in \mathbb{R}$  et  $h$  est la largeur de bande. Alors, pour tous  $f \in \mathbb{L}^2(\mathbb{R}, d\nu)$  nous avons  $\mathbb{E}(T_K) = \langle k_h * f, f \rangle$ , où  $*$  est le produit de convolution par rapport à la mesure  $\nu$  et  $k_h(u) = \frac{1}{h} k\left(\frac{u}{h}\right)$ ,  $\forall u \in \mathbb{R}$ . Ainsi, lorsque la bande  $h$  est bien choisie,  $T_K$  peut être considéré comme un estimateur pertinent de  $\|f\|^2$ .

À partir des choix des deux exemples ci-dessus pour  $K$ , nous avons vu que le test d'hypothèse  $V_K$  peut être considéré comme un estimateur pertinent de  $\|f\|^2$ . Ainsi, il semble être raisonnable de considérer un test qui rejette  $(H_0)$  quand  $V_K$  est "assez grand".

Nous avons défini les valeurs critiques utilisées dans nos tests en définissant,

$$V_K^{(0)} = \frac{\frac{1}{n(n-1)} \sum_{i \neq j=1}^n K(X_i, X_j) \epsilon_i \epsilon_j}{\frac{1}{n} \sum_{i=1}^{n/2} (\epsilon'_{2i-1} - \epsilon'_{2i})^2}. \quad (1.3.14)$$

Pour  $\alpha$  dans  $(0, 1)$ , on a noté la  $q_{K,1-\alpha}^{(X)}$  le  $(1 - \alpha)$ -quantile de la distribution de  $V_K^{(0)}$  conditionnellement à  $X$ . Nous rejetons  $(H_0)$  quand  $V_K > q_{K,1-\alpha}^{(X)}$ . La fonction de test correspondante est définie par

$$\Phi_{K,\alpha} = \mathbb{1}\{V_K > q_{K,1-\alpha}^{(X)}\}. \quad (1.3.15)$$

### Probabilités d'erreurs de type I et II

Sous  $(H_0)$ ,  $V_K$  et  $V_K^{(0)}$  ont la même distribution conditionnellement à  $X$ , pour tous  $\alpha \in (0, 1)$ , nous avons

$$\mathbb{P}_{(H_0)} \left( V_K > q_{K,1-\alpha}^{(X)} \middle| X \right) \leq \alpha.$$

En prenant l'espérance par rapport à  $X$ , nous obtenons

$$\mathbb{P}_{(H_0)} (\Phi_{K,\alpha} = 1) \leq \alpha.$$

En notant  $q_{K,1-\beta/2}^\alpha$  le  $(1 - \beta/2)$ -quantile du quantile conditionnel  $q_{K,1-\alpha}^{(X)}$ , nous avons donné la condition pour minorer l'erreur de seconde espèce

**Proposition 1.2.** *Soit  $\alpha, \beta$  les niveaux fixés dans  $(0, 1)$ . Nous avons*

$$\mathbb{P}_f \left( V_K \leq q_{K,1-\beta/2}^\alpha \right) \leq \beta/2,$$

*sous réserve que*

$$\langle K[f], f \rangle \geq \sqrt{\frac{16A_K + 8B_K}{\beta}} + D_{n,\beta} q_{K,1-\beta/2}^\alpha, \quad (1.3.16)$$

avec

$$\begin{aligned} A_K &= \frac{n-2}{n(n-1)} \int_E (K[f](x))^2 [f^2(x) + \sigma^2] d\nu(x), \\ B_K &= \frac{1}{n(n-1)} \int_{E^2} K^2(x, y) [f^2(x) + \sigma^2] [f^2(y) + \sigma^2] d\nu(x)d\nu(y), \\ D_{n,\beta} &= \sigma^2 + a^2 + \frac{4\sigma^2}{n} \sqrt{\left(\frac{n}{2} + \frac{na^2}{\sigma^2}\right) \ln\left(\frac{2}{\beta}\right) + \frac{4\sigma^2}{n} \ln\left(\frac{2}{\beta}\right)}. \end{aligned}$$

Ainsi nous avons, sous (1.3.16),

$$\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta.$$

De plus, il existe une constante  $\kappa > 0$  telle que, pour chaque  $K$  et  $n \geq 32 \ln(2/\alpha)$ ,

$$q_{K,1-\beta/2}^\alpha \leq \frac{2\kappa}{\sqrt{n(n-1)}} \ln\left(\frac{2}{\alpha}\right) \sqrt{\frac{2 \int_{E^2} K^2(x, y) d\nu(x)d\nu(y)}{\beta}}. \quad (1.3.17)$$

Nous avons également donné une condition sur  $f$  pour que le test soit puissant.

**Théorème 1.3.** Soit  $\alpha, \beta$  les niveaux fixés dans  $(0, 1)$ ,  $\kappa$  une constante positive,  $K$  une fonction de noyau symétrique, et  $\Phi_{K,\alpha}$  le test défini par (1.3.15). Soit  $C_K$  une limite supérieure pour  $\int_{E^2} K^2(x, y) d\nu(x)d\nu(y)$ . Pour tout  $n \geq 32 \ln(2/\alpha)$ , nous avons  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$ , sous réserve que

$$\begin{aligned} \|f\|^2 &\geq \|f - K[f]\|^2 + \frac{16(\|f\|_\infty^2 + \sigma^2)}{n\beta} \\ &+ \frac{4}{\sqrt{n(n-1)}\beta} \left( \kappa D_{n,\beta} \ln\left(\frac{2}{\alpha}\right) + \sqrt{2}(\|f\|_\infty^2 + \sigma^2) \right) \sqrt{C_K}. \end{aligned} \quad (1.3.18)$$

Par exemple,  $C_K$  peut être pris comme suit.

1.  $C_K = D$  quand  $K$  est choisi comme dans l'Exemple 1, en considérant une base orthonormée  $\{\phi_\lambda, \lambda \in \Lambda\}$  d'un sous-espace  $D$ -dimensionnel  $S$  de  $\mathbb{L}^2([0, 1], d\nu)$ .
2.  $C_K = \frac{\|\nu\|_\infty}{h\sqrt{2\pi}}$  quand  $K$  est choisi comme dans l'Exemple 2.

### Noyaux de projection.

Considérons le cas particulier où le noyau  $K$  est le noyau de projection de l'Exemple 1 sur l'espace engendré par les fonctions de la base de Haar définies ci-dessus. Nous visons à évaluer les taux de séparation uniformes sur les espaces de Besov.

Soit  $\{\phi_0, \phi_{(j,k)}, j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}\}$  la base de Haar  $\mathbb{L}^2([0, 1])$  avec

$$\phi_0(x) = \mathbb{1}_{[0,1]}(x) \quad \text{et} \quad \phi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \quad (1.3.19)$$

où  $\psi(x) = \mathbb{1}_{[0,1/2)}(x) - \mathbb{1}_{[1/2,1]}(x)$ . Le sous-espace linéaire  $S$  est engendré par un sous-ensemble de la base de Haar. Plus précisément, on note  $S_0$  le sous-espace de  $\mathbb{L}^2([0,1])$  engendré par  $\phi_0$ , et on définit

$$K_0(x, x') = \phi_0(x)\phi_0(x'). \quad (1.3.20)$$

Nous considérons également, pour  $J \geq 1$ , le sous-espace  $S_J$  engendré par  $\{\phi_\lambda, \lambda \in \{0\} \cup \Lambda_J\}$  avec  $\Lambda_J = \{(j, k), j \in \{0, \dots, J-1\}, k \in \{0, \dots, 2^j-1\}\}$ , et

$$K_J(x, x') = \sum_{\lambda \in \{0\} \cup \Lambda_J} \phi_\lambda(x)\phi_\lambda(x'). \quad (1.3.21)$$

Nous fixons  $\alpha_0 = \langle f, \phi_0 \rangle$  et pour tout  $j \in \mathbb{N}$ ,  $k \in \{0, \dots, 2^j-1\}$ ,  $\alpha_{j,k} = \langle f, \phi_{j,k} \rangle$ . Nous présentons maintenant le corps de Besov défini pour  $\delta > 0$ ,  $R > 0$  par

$$\mathcal{B}_{2,\infty}^\delta(R) = \left\{ f \in \mathbb{L}^2([0,1], d\nu), f = \alpha_0\phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha_{j,k}\phi_{j,k} / \alpha_0^2 \leq R^2, \forall j \in \mathbb{N}, \sum_{k=0}^{2^j-1} \alpha_{j,k}^2 \leq R^2 2^{-2j\delta} \right\}.$$

Pour tout  $J \geq 0$ , nous considérons la fonction noyau  $K_J$  définie par (1.3.20), (1.3.21) et la fonction de test associée  $\Phi_{K_J, \alpha}$  définie dans (1.3.15) avec  $K = K_J$ . Nous donnons une condition sur  $\|f\|^2$  pour  $f \in \mathcal{B}_{2,\infty}^\delta(R)$  qui garantit que la puissance de notre test est supérieure à  $1 - \beta$ .

**Proposition 1.3.** *Soit  $\alpha, \beta \in (0,1)$ . Pour tout  $J \geq 0$ , on considère  $K_J$  définie par (1.3.20), (1.3.21) et la fonction test  $\Phi_{K_{J^*}, \alpha} = \mathbb{1}\{V_{K_{J^*}} > q_{K_{J^*}, 1-\alpha}^{(X)}\}$  où*

$$J^* = \left\lceil \log_2 \left( n^{2/(1+4\delta)} \right) \right\rceil. \quad (1.3.22)$$

Pour tout  $f \in \mathcal{B}_{2,\infty}^\delta(R)$  tel que

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, \|f\|_\infty) n^{-4\delta/(1+4\delta)}, \quad (1.3.23)$$

on a  $\mathbb{P}_f(\Phi_{K_{J^*}, \alpha} = 0) \leq \beta$ .

*Commentaires.*

1. Les limites inférieures non asymptotiques des taux de test en détection de signal sur des espaces de Besov sont données dans [Baraud et al. \(2002\)](#). Ces limites inférieures coïncident avec les limites indiquées dans (1.3.23). Notre résultat est plus optimal dans ce cas.
2. Dans (1.3.22),  $J^*$  dépend de  $\delta$ , le paramètre de régularité de l'espace de Besov, ce qui conduit à la question naturelle du choix de ce paramètre. Afin de proposer une procédure adaptative par rapport à la régularité de la fonction de régression inconnue  $f$ , nous introduisons des tests agrégés.

## Noyaux Gaussiens

Considérons le cas particulier où le noyau  $K$  est le noyau Gaussien de l'*Exemple 2*. Nous visons à évaluer les taux de séparation uniformes sur les espaces de Sobolev. Soit  $E = \mathbb{R}$  et  $\mathcal{L} = \mathbb{N}^*$ . Pour  $x, y$  dans  $\mathbb{R}$  et  $h = 2^{-l}$ , pour tout  $l \in \mathcal{L}$ , nous considérons

$$K_l(x, y) = \frac{1}{2^{-l}} k\left(\frac{x-y}{2^{-l}}\right), \quad (1.3.24)$$

avec

$$k(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right).$$

Pour  $\delta > 0$  on considère l'espace de Sobolev  $\mathcal{S}_\delta(R)$  défini par

$$\mathcal{S}_\delta(R) = \left\{ s : \mathbb{R} \rightarrow \mathbb{R} \mid s \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |u|^{2\delta} |\hat{s}(u)|^2 du \leq 2\pi R^2 \right\},$$

où  $\hat{s}$  dénote la transformée de Fourier de  $s$ :  $\hat{s}(u) = \int_{\mathbb{R}} s(x) e^{i(x,u)} dx$ .

Pour tout  $l \in \mathcal{L}$ , nous considérons la fonction noyau  $K_l$  définie par (1.3.24) et la fonction de test associée  $\Phi_{K_l, \alpha}$  définie par (1.3.15) avec  $K = K_l$ . Nous donnons une condition sur  $f \in \mathcal{S}_\delta(R)$  qui assure que la puissance de notre test est supérieure à  $1 - \beta$ .

**Proposition 1.4.** *Soit  $\alpha, \beta \in (0, 1)$ . Pour tout  $l \in \mathcal{L}$ , soit  $K_l$  définie par (1.3.24) et la fonction test  $\Phi_{K_l, \alpha} = \mathbb{1}\{V_{K_l} > q_{K_l, 1-\alpha}^{(X)}\}$ . Nous fixons*

$$l^* = \left\lceil \log_2 \left( n^{\frac{2}{1+4\delta}} \right) \right\rceil. \quad (1.3.25)$$

Pour tout  $f \in \mathcal{S}_\delta(R)$  tel que

$$\|f\|^2 \geq C(\alpha, \beta, \kappa, \sigma, \|f\|_\infty) n^{-4\delta/(1+4\delta)}, \quad (1.3.26)$$

on a  $\mathbb{P}_f(\Phi_{K_{l^*}, \alpha} = 0) \leq \beta$ .

*Commentaires.*

1. Comme dans la proposition 1.3, la condition de minoration de (1.3.26) est une borne classique pour les taux de séparation des tests par rapport aux classes régulières d'alternatives telles que les espaces de fonctions höldériennes (voir [Ingster \(1993\)](#)) pour les taux de test minimax non paramétriques dans diverses configurations.
2. Les limites inférieures non asymptotiques des taux de test en détection de signal sur des espaces de Sobolev sont données dans [Fromont and Lévy-Leduc \(2006\)](#). Ces limites coïncident avec les limites données dans (1.3.26).
3. Dans (1.3.25), comme précédemment,  $l^*$  dépend de  $\delta$ , le paramètre de régularité de l'espace de Sobolev, cela conduit donc à la question naturelle du choix de ce paramètre traité par les tests agrégés.

### Tests agrégés basés sur des collections de fonctions noyau

Pour éviter de choisir le noyau et ses paramètres dans les tests simples, nous avons considéré une collection de fonctions noyau au lieu d'une seule. Nous avons défini une procédure de test multiple en agrégeant les tests individuels correspondant, avec un choix adapté des valeurs critiques.

On introduit une collection finie  $\{K_m, m \in \mathcal{M}\}$  de fonctions noyau symétriques définies sur:  $E \times E$  à valeur dans  $\mathbb{R}$ . Pour  $m \in \mathcal{M}$ , on remplace  $K$  dans (1.3.8) et (1.3.14) par  $K_m$  pour définir  $V_{K_m}$  et  $V_{K_m}^{(0)}$ . On pose  $\{w_m, m \in \mathcal{M}\}$  une collection de nombres positifs tels que  $\sum_{m \in \mathcal{M}} e^{-w_m} \leq 1$ . Conditionnellement sur  $X$ , pour  $u \in (0, 1)$ , nous notons par  $q_{m,1-u}^{(X)}$  le  $(1-u)$ -quantile de  $V_{K_m}^{(0)}$ . Etant donné  $\alpha$  dans  $(0, 1)$ , nous considérons le test qui rejette ( $H_0$ ) quand il existe au moins un  $m$  dans  $\mathcal{M}$  tel que

$$V_{K_m} > q_{m,1-u_\alpha^{(X)}}^{(X)} e^{-w_m},$$

où  $u_\alpha^{(X)}$  est défini par

$$u_\alpha^{(X)} = \sup \left\{ u > 0, \mathbb{P} \left( \sup_{m \in \mathcal{M}} \left( V_{K_m} - q_{m,1-ue^{-w_m}}^{(X)} \right) > 0 \middle| X \right) \leq \alpha \right\}. \quad (1.3.27)$$

Nous avons considéré la fonction test  $\Phi_\alpha$  définie par

$$\Phi_\alpha = \mathbb{1} \left\{ \sup_{m \in \mathcal{M}} \left( V_{K_m} - q_{m,1-u_\alpha^{(X)}}^{(X)} e^{-w_m} \right) > 0 \right\}. \quad (1.3.28)$$

Nous avons fourni un contrôle des erreurs de 1ère et de 2ème espèce pour le test  $\Phi_\alpha$ .

**Théorème 1.4.** *Soit  $\alpha, \beta$  des niveaux fixés dans  $(0, 1)$  et  $\Phi_\alpha$  le test défini par (1.3.28). Nous avons*

$$\mathbb{P}_{(H_0)} (\Phi_\alpha = 1) \leq \alpha. \quad (1.3.29)$$

*Et pour toute fonction de régression  $f$ , nous avons*

$$\mathbb{P}_f (\Phi_\alpha = 0) \leq \beta, \quad (1.3.30)$$

*lorsqu'il existe  $m$  dans  $\mathcal{M}$  telle que*

$$\mathbb{P} \left( V_{K_m} \leq q_{K_m,1-\alpha e^{-w_m}}^{(X)} \right) \leq \beta.$$

*Commentaires.* Ce théorème montre que le test agrégé est de niveau  $\alpha$ , pour tout  $n$ . De plus, dès que l'erreur de 2ème espèce est contrôlée par  $\beta$  pour au moins un test de la collection, il en va de même pour la procédure agrégée avec le prix que le niveau  $\alpha$  est remplacé par  $\alpha e^{-w_m}$  pour garantir que la procédure agrégée est de niveau  $\alpha$ .

### Cas d'une collection de noyaux de projection

Spécifions les performances du test agrégé pour une collection de noyaux de projection.

**Corollary 1.5.** *Soit  $\alpha, \beta$  des niveaux fixés dans  $(0, 1)$ . Soit  $\{S_m, m \in \mathcal{M}\}$  une collection finie de sous-espaces linéaires de  $\mathbb{L}^2([0, 1], d\nu)$ , engendrés par les fonctions  $\{\phi_\lambda, \lambda \in \Lambda_m\}$ . Nous supposons que la dimension de  $S_m$  est égale à  $D_m$ . Nous posons, pour tout  $m \in \mathcal{M}$ ,  $K_m(x, y) = \sum_{\lambda \in \Lambda_m} \phi_\lambda(x)\phi_\lambda(y)$ . Soit  $\Phi_\alpha$  défini par (1.3.28) avec la collection de noyaux  $\{K_m, m \in \mathcal{M}\}$  et une collection  $\{w_m, m \in \mathcal{M}\}$  des nombres positifs tels que  $\sum_{m \in \mathcal{M}} e^{-w_m} \leq 1$ .*

Alors  $\Phi_\alpha$  est un test de niveau  $\alpha$ . En outre,  $\mathbb{P}_f(\Phi_\alpha = 0) \leq \beta$  si

$$\|f\|^2 \geq \inf_{m \in \mathcal{M}} \left\{ \|f - \Pi_{S_m}(f)\|^2 + \frac{16(\|f\|_\infty^2 + \sigma^2)}{n\beta} + \frac{4\sqrt{D_m}}{\sqrt{n(n-1)\beta}} \left( \kappa D_{n,\beta} \left( \ln\left(\frac{2}{\alpha}\right) + w_m \right) + \sqrt{2}(\|f\|_\infty^2 + \sigma^2) \right) \right\}, \quad (1.3.31)$$

où  $\kappa > 0$  et  $n \geq 32 \ln(\alpha/2)$ .

*Commentaires.* En comparant ce résultat avec celui obtenu dans le Théorème 1.3 pour le test simple basé sur un noyau de projection, nous pouvons voir que la procédure de test multiple permet d'obtenir l'infimum sur tout  $m$  dans  $\mathcal{M}$  dans le membre droit de (1.3.31) au prix du terme supplémentaire  $w_m$ .

Considérons maintenant le cas particulier où  $\mathcal{M} = \mathcal{M}_{\bar{J}} = \{J, 0 \leq J \leq \bar{J}\}$  pour un certain  $\bar{J} \geq 1$  et  $w_J = 2(\ln(J+1) + \ln(\pi/\sqrt{6}))$ .

Nous considérons  $\Phi_\alpha^{(1)}$ , le test défini par (1.3.28) avec la collection de noyaux  $\{K_J, 0 \leq J \leq \bar{J}\}$  où  $K_0, K_J, 0 < J \leq \bar{J}$  défini dans (1.3.20), (1.3.21). Nous obtenons du Corollaire 1.5 qu'il existe une constante  $C(\alpha, \beta, \sigma, \|f\|_\infty)$  tel que  $\mathbb{P}_f(\Phi_\alpha^{(1)} = 0) \leq \beta$  sous réserve que

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, \|f\|_\infty) \inf_{J \in \mathcal{M}_{\bar{J}}} \left\{ \|f - \Pi_{S_J}(f)\|^2 + \ln(J+2) \frac{2^{J/2}}{n} \right\}. \quad (1.3.32)$$

Pour tout  $\delta > 0, R, R' > 0$  nous considérons

$$\mathcal{B}_{2,\infty}^\delta(R, R') = \left\{ f : f \in \mathcal{B}_{2,\infty}^\delta(R), \|f\|_\infty \leq R' \right\}. \quad (1.3.33)$$

**Corollary 1.6.** *Fixons  $\alpha, \beta \in (0, 1)$ . Pour tout  $J \in \mathcal{M}_{\bar{J}}$ , nous considérons la fonction de test  $\Phi_\alpha^{(1)}$ . En choisissant  $n$  tel que  $\ln \ln(n) \geq 1$ ,  $2^J \geq n^2$ . Ensuite, pour tout  $\delta, R, R' > 0$  nous fixons*

$$J^{**} = \left\lceil \log_2 \left( \left( \frac{n}{\ln \ln(n)} \right)^{\frac{2}{4\delta+1}} \right) \right\rceil.$$



Pour tout  $f \in \mathcal{B}_{2,\infty}^\delta(R, R')$  tel que

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, R') \left( \frac{\ln \ln(n)}{n} \right)^{\frac{4\delta}{4\delta+1}}, \quad (1.3.34)$$

on a  $\mathbb{P}_f(\Phi_\alpha^{(1)} = 0) \leq \beta$ .

*Commentaire.* Nous obtenons une minoration dans (1.3.34) de l'ordre de  $(\ln \ln(n)/n)^{4\delta/(1+4\delta)}$ . Ce taux de test s'est avéré optimal pour la détection du signal dans un bruit blanc gaussien par [Spokoiny et al. \(1996\)](#). En particulier, il a montré qu'un facteur logarithmique est le prix à payer pour l'adaptation.

### Cas d'une collection de noyaux Gaussiens

Nous considérons ici le test agrégé basé sur une collection de noyaux gaussiens.

**Corollaire 1.7.** Soient  $\alpha, \beta \in (0, 1)$ ,  $\{h_l, l \in \mathcal{L}\}$  une collection de fenêtres, nous considérons  $\{K_l, l \in \mathcal{L}\}$  une collection de noyaux gaussiens correspondant à la collection ci-dessus où  $K_l$  défini dans (1.3.24). Soit  $\Phi_\alpha$  défini par (1.3.28) avec la collection de noyaux  $\{K_l, l \in \mathcal{L}\}$  et une collection  $\{w_l, l \in \mathcal{L}\}$  de nombres positifs tels que  $\sum_{l \in \mathcal{L}} e^{-w_l} \leq 1$ .

Alors  $\Phi_\alpha$  est un test de niveau  $\alpha$ . De plus, il existe  $\kappa > 0$  tel que si

$$\|f\|^2 \geq \inf_{l \in \mathcal{L}} \left\{ \|f - k_l * f\|^2 + \frac{16(\|f\|_\infty^2 + \sigma^2)}{n\beta} + \frac{4\|\nu\|_\infty}{(2\pi)^{1/4}\sqrt{n(n-1)\beta h_l}} \left( \kappa D_{n,\beta} \left( \ln \left( \frac{2}{\alpha} \right) + w_l \right) + \sqrt{2}(\|f\|_\infty^2 + \sigma^2) \right) \right\}, \quad (1.3.35)$$

nous obtenons  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$ .

Nous considérons le cas particulier où nous prenons  $\mathcal{L} = \mathbb{N}^*$ ,  $h_l = 2^{-l}$  et  $w_l = 2(\ln(l+1) + \ln(\pi^2/6))$  pour tout  $l \in \mathcal{L}$ . Soit  $\Phi_\alpha^{(2)}$  le test défini par (1.3.28) avec la collection de noyaux gaussiens  $\{K_l, l \in \mathcal{L}\}$  et  $\{w_l, l \in \mathcal{L}\}$ . Nous obtenons du Corollaire 1.7 qu'il existe  $C(\alpha, \beta, \sigma, \|f\|_\infty)$  tel que  $\mathbb{P}_f(\Phi_\alpha^{(2)} = 0) \leq \beta$  sous réserve que

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, \|f\|_\infty) \inf_{l \in \mathcal{L}} \left\{ \|f - k_l * f\|^2 + \frac{w_l}{n\sqrt{2^{-l}}} \right\}. \quad (1.3.36)$$

Pour  $\delta > 0, R, R' > 0$  nous considérons

$$\mathcal{S}_\delta(R, R') = \left\{ f : f \in \mathcal{S}_\delta(R), \|f\|_\infty \leq R' \right\}. \quad (1.3.37)$$

**Corollary 1.8.** *Soient  $\alpha, \beta \in (0, 1)$ . Pour tout  $l \in \mathcal{L}$ , nous considérons la fonction test  $\Phi_\alpha^{(2)}$  et supposons que  $\ln \ln n \geq 1$ . Pour tout  $\delta > 0, R, R' > 0$ , nous fixons*

$$l^{**} = \left\lceil \log_2 \left( \left( \frac{n}{\ln \ln(n)} \right)^{\frac{2}{4\delta+1}} \right) \right\rceil.$$

Pour tout  $f \in \mathcal{S}_\delta(R, R')$  tel que

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, R') \left( \frac{\ln \ln(n)}{n} \right)^{\frac{4\delta}{4\delta+1}},$$

on a  $\mathbb{P}_f(\Phi_\alpha^{(2)} = 0) \leq \beta$ .

*Commentaires.* Le taux de test est de l'ordre de  $(\ln \ln(n)/n)^{4\delta/(1+4\delta)}$ . Il a été démontré dans [Castillo, Lévy-Leduc, and Matias \(2006\)](#) que ce taux était optimal sur les boules de Sobolev périodiques jusqu'au logarithme.

À la fin de la seconde partie, nous avons effectué des tests numériques pour évaluer la qualité probabilité de commettre une erreur de 1ère espèce et la probabilité de commettre une erreur de 2ème espèce de trois procédures de tests agrégés multiples. La première  $\Phi_\alpha^{(1)}$  est basée sur la collection de fonctions noyaux de projection, la deuxième  $\Phi_\alpha^{(2)}$  est basée sur la collection de fonctions noyaux Gaussiens et la troisième  $\Phi_\alpha^{(3)}$  utilise un mélange des deux précédentes collections (projection et gaussiens). Les résultats de ces expériences numériques ont montré que les tests multiples répondent à des propriétés non asymptotiques et puissantes.



# Part I

## Part 1: Gaussian kernel regression



# Chapitre 2

## Distribution regression model with a Reproducing Kernel Hilbert Space approach

The results of this chapter are joint work with Loubes, J-M., Risser, L. and Balaresque, P. [Bui et al. \(2018\)](#).

### Contents

---

<b>2.1</b>	<b>Introduction</b>	<b>34</b>
<b>2.2</b>	<b>Preliminaries</b>	<b>35</b>
2.2.1	Wasserstein space and barycenter of a probability in Wasserstein space	35
2.2.2	Reproducing kernel Hilbert space	37
<b>2.3</b>	<b>Kernel on Wasserstein space <math>\mathcal{W}_2(\mathbb{R})</math></b>	<b>38</b>
2.3.1	The Wasserstein space on $\mathbb{R}$	38
2.3.2	Kernel	40
<b>2.4</b>	<b>Regression</b>	<b>41</b>
2.4.1	Setting	41
2.4.2	Universal kernel	42
<b>2.5</b>	<b>Numerical Simulations and Real data application</b>	<b>46</b>
2.5.1	Simulation in the regression model with one-dimensional distribution input	46
2.5.2	Simulation in the regression model with two-dimensional distribution input	52
2.5.3	Application on evolution of hearing sensitivity	55
<b>2.6</b>	<b>Discussion</b>	<b>59</b>

---

We introduce a new distribution regression model for probability distributions. This model is based on a Reproducing Kernel Hilbert Space (RKHS) regression framework, where universal kernels are built using Wasserstein distances for distributions belonging to  $\mathcal{W}_2(\Omega)$  and  $\Omega$  is a compact subspace of  $\mathbb{R}$ . We prove the universal kernel property of such kernels and use this setting to perform regressions on functions. Different regression models are first compared with the proposed one on simulated functional data for both one-dimensional and two-dimensional distributions. We then apply our regression model to transient evoked otoacoustic emission (TEOAE) distribution responses and real predictors of the age.

## 2.1 Introduction

Regression analysis is a predictive modeling technique that has been widely studied over the last decades with the goal to investigate relationships between predictors and responses (inputs and outputs) in regression models, see for instance [Kutner et al. \(2004\)](#), [Azaïs \(2006\)](#) and references therein. When the inputs belong to functional spaces, different strategies have been investigated and used in several application domains about functional data analysis [Neter et al. \(1996\)](#), [Ramsay and Silverman \(2007\)](#). The Reproducing Kernel Hilbert Space (RKHS) framework became recently popular to extend the results of the statistical learning theory in the context of regression of functional data as well as to develop estimation procedures of functional valued functions  $f$  as in [Kadri et al. \(2010\)](#); [Preda \(2007\)](#) for instance. This framework is particularly important in the field of statistical learning theory because of the so-called *Representer theorem*, which states that every function can be written as a linear combination of the kernel function evaluated at training points [Berlinet and Thomas-Agnan \(2011\)](#), providing a setting to estimate functions over a wide range of functional spaces.

In our framework, we aim to solve the regression problem with inputs belonging to probability distribution spaces whose outputs are real values. Specially, we consider the model

$$y_i = f(\mu_i) + \sigma\epsilon_i, \quad (2.1.1)$$

where  $\{\mu_i\}_{i=1}^n$  are probability distributions on  $\mathbb{R}$ ,  $\{y_i\}_{i=1}^n$  are real numbers and the  $\epsilon_i$  represent an independent and identically distributed Gaussian noise and  $\sigma$  is the level of noise. As in classical regression models, this setting estimates an unknown function  $f$  from the observations  $\{(\mu_i, y_i)\}_{i=1}^n$ .

Distribution regression problem has become a major concern in the recent years. A common approach is to look at embedding into a Hilbert space using kernels as for instance in [Smola et al. \(2007\)](#) with the mean embedding or see in [Muandet et al. \(2012\)](#) for a review. The studies for regression with distribution inputs and real-valued responses are one of the most popular research topics in this field, for instance [Póczos et al. \(2013\)](#) and [Oliva et al. \(2014\)](#). For more information about distribution regression theoretical properties, we refer to [Szabó, Sriperumbudur,](#)

Póczos, and Gretton (2016) and Muandet, Fukumizu, Sriperumbudur, Schölkopf, et al. (2017).

In what follows, we will consider kernels built using the Wasserstein distance. (see in Villani (2008)). Some kernels with this metric have been developed in Kolouri et al. (2015); Peyré, Cuturi, and Solomon (2016). In studies of Kolouri et al. (2015), authors built a sliced Wasserstein kernel based on Wasserstein distance, which was proved to be a positive definite kernel for absolutely continuous distributions. More generally, for any distributions, authors in Bachoc, Gamboa, Loubes, and Venet (2017) built a family of positive definite kernel based also on Wasserstein distance for distributions on the real line. The properties of characteristic kernels were explored further in Sriperumbudur, Gretton, Fukumizu, Lanckriet, and Schölkopf (2008); Sriperumbudur, Gretton, Fukumizu, Schölkopf, and Lanckriet (2010b); Sriperumbudur, Fukumizu, and Lanckriet (2011). Besides that, a closely related to these kernel for achieving the richness of RKHS was previously studied through the so-called universal kernels in Steinwart (2001). Universality property for kernels is stronger than the characteristic property. Actually, using results in Sriperumbudur, Fukumizu, and Lanckriet (2010a) all universal kernels are characteristic. Within the framework of Christmann and Steinwart (2010), the authors prove universality of Gaussian-type RBF-kernel. Further, combining the positive definite kernel built by Bachoc, Gamboa, Loubes, and Venet (2017) with the Gaussian-type RBF-kernel in Christmann and Steinwart (2010), we consider a universal kernel based on the Wasserstein distance between two probability distributions on a compact space.

The paper falls into the following parts. In Section 2.3, we first recall important concepts about kernels on Wasserstein spaces  $\mathcal{W}_2(\mathbb{R})$ . Section 2.4 deals with the proposed setting of distribution regression models. We then assess the numerical performance of this method in Section 2.5. The tests are first performed on simulated generated data in one and two dimensions to compare our model with state-of-the-art ones. Then, we study the relationship between the age and hearing sensitivity by using TEOAEs recording that are acquired by stimulating with a very short but strong broadband stimulus. These recordings are then the ear responses by emitting a sound track on a given frequency. More precisely, we predict the age of the subjects, on which they were acquired using the proposed distribution regression model, from TEOAE data. Discussions are finally drawn Section 2.6.

## 2.2 Preliminaries

In this section, we recall some basis definitions and properties of the Wasserstein space and reproducing kernel Hilbert space.

### 2.2.1 Wasserstein space and barycenter of a probability in Wasserstein space

The following is taken mainly from Gouic and Loubes (2016).



**Wasserstein space**

Given two points  $x, y$  in a metric space  $(E, d)$  their mid-point is the point  $z \in E$  such that

$$d(x, z) = d(z, y) = \frac{1}{2}d(x, y).$$

**Definition 2.1. (Geodesic space)** *A space  $(E, d)$  is called a geodesic space if*

1.  $(E, d)$  is a complete metric space and
2. every two points  $x, y \in E$  have a mid-point  $z \in E$ .

Set  $p \geq 1$  and let  $(E, d)$  be a geodesic space and  $\mu$  a probability measure on  $(E, d)$  such that

$$\int d^p(x, x_0) d\mu(x) < \infty, \quad (2.2.1)$$

for some (and thus any)  $x_0 \in E$ .

The set of all probability measures satisfying (2.2.1) is denoted  $\mathcal{W}_p(E)$ .

**Definition 2.2. (Wasserstein space)** *Set  $p \geq 1$  and let  $(E, d)$  be a metric space. Given two measures  $\mu, \nu \in \mathcal{W}_p(E)$ , we denote by  $\Pi(\mu, \nu)$  the set of all probability measures  $\pi$  over the product set  $E \times E$  with first, (resp. second) marginal  $\mu$  (resp.  $\nu$ ). The transportation cost with a cost function  $d^p$  between two measures  $\mu, \nu$  in  $\mathcal{W}_p(E)$ , is defined as*

$$\mathcal{T}_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int d^p(x, y) d\pi. \quad (2.2.2)$$

The transportation cost allows to endow the set  $\mathcal{W}_p(E)$  with a metric  $W_p$  defined by

$$W_p(\mu, \nu) = \mathcal{T}_p(\mu, \nu)^{1/p}. \quad (2.2.3)$$

This metric is known as the  $p$ -Wasserstein distance and the metric space  $(\mathcal{W}_p(E), W_p)$  is called the Wasserstein space of  $(E, d)$ .

A probability measure  $\pi$  in  $\Pi(\mu, \nu)$  performing the infimum in (2.2.2) is called an optimal coupling. This vocabulary transfers to a random vector  $(X_1, X_2)$  with distribution  $\pi$ .

The Wasserstein space  $\mathcal{W}_p(E)$  inherits several properties of the base space  $E$ , find more in [Villani \(2008\)](#).

**Theorem 2.1. (Topology of the Wasserstein space)** *Let  $E$  be a complete separable metric space and  $p \geq 1$ . Then the Wasserstein space  $\mathcal{W}_p(E)$ , metrized by the Wasserstein distance  $W_p$ , is also a complete separable metric space.*

## Barycenters of a probability in Wasserstein space

**Definition 2.3. (Barycenter)** Set  $p \geq 1$  and let  $(E, d)$  be a geodesic space and  $\mu \in \mathcal{W}_p(E)$ . A point  $x_0 \in E$  is called a  $p$ -barycenter of  $\mu$  if

$$\int d^p(x, x_0) d\mu(x) = \inf \left\{ \int d^p(x, y) d\mu(x); y \in E \right\}$$

Barycenter do not always exists. The following proposition ensure the existence of a barycenter.

**Proposition 2.2.** Set  $p \geq 1$  and let  $(E, d)$  be a locally compact geodesic space and  $\mu \in \mathcal{W}_p(E)$ . Then, there exists a barycenter of  $\mu$ .

Considering a random probability measure  $\tilde{\mu}$  in  $\mathcal{W}_p(E)$ , following a distribution  $\mathbb{P}$ . This probability  $\mathbb{P}$  is chosen in the space  $\mathcal{W}_p(\mathcal{W}_p(E))$  endowed with metric  $W_p$ . Note that we use the same notation Wasserstein distance over  $\mathcal{W}_p(\mathcal{W}_p(E))$  and  $\mathcal{W}_p(E)$ . This if  $\tilde{\mu} \in \mathcal{W}_p(E)$  is a-random measure with distribution  $\mathbb{P}$ , then for all  $\nu \in \mathcal{W}_p(E)$  we can write

$$W_p^p(\delta_\nu, \mathbb{P}) = \mathbb{E} \left( W_p^p(\nu, \tilde{\mu}) \right) = \int W_p^p(\nu, \mu) d\mathbb{P}(\mu).$$

For a probability  $\mathcal{W}_p(\mathcal{W}_p(E))$ , consider a minimizer over  $\nu \in \mathcal{W}_p(E)$  of

$$\nu \mapsto \mathbb{E} \left[ W_p^p(\nu, \tilde{\mu}) \right] = W_p^p(\delta_\nu, \mathbb{P}),$$

where  $\tilde{\mu}$  is a random probability of  $\mathcal{W}_p(E)$  with distribution  $\mathbb{P}$ . If exists, this probability measure is a barycenter of  $\mathbb{P}$ .

**Theorem 2.3. (Existence of a Wasserstein barycenter)** Set  $p \geq 1$  and let  $(E, d)$  be a locally compact geodesic space. Hence, for  $\mathbb{P} \in \mathcal{W}_p(\mathcal{W}_p(E))$ , there exists a barycenter  $\bar{\mu}_{\mathbb{P}}$  defined as

$$\bar{\mu}_{\mathbb{P}} \in \arg \min_{\nu \in \mathcal{W}_p(E)} \mathbb{E} \left[ W_p^p(\nu, \tilde{\mu}) \right],$$

for  $\tilde{\mu}$  be a random measure with distribution  $\mathbb{P}$ .

## 2.2.2 Reproducing kernel Hilbert space

### Construction of kernels

The following is taken mainly from [Steinwart and Christmann \(2008\)](#).

**Definition 2.4. (Inner product)** Let  $\mathcal{H}$  be a vector space over  $\mathbb{R}$ . A function  $\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$  is said to be an inner product on  $\mathcal{H}$  if

1.  $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle_{\mathcal{H}} = \alpha_1 \langle f_1, g \rangle_{\mathcal{H}} + \alpha_2 \langle f_2, g \rangle_{\mathcal{H}}$ ,

2.  $\langle f, g \rangle_{\mathcal{H}} = \langle g, f \rangle_{\mathcal{H}}$ ,
3.  $\langle f, f \rangle_{\mathcal{H}} \leq 0$  and  $\langle f, f \rangle_{\mathcal{H}} = 0$  if and only if  $f = 0$ .

We can define a norm using the inner product as  $\|f\|_{\mathcal{H}} := \sqrt{\langle f, f \rangle_{\mathcal{H}}}$ .

A Hilbert space is a space on which an inner product is defined, along with an additional technical condition. We now define a kernel.

**Definition 2.5. (Kernel)** Let  $\mathcal{X}$  be a non-empty set. A function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is called a kernel if there exists an  $\mathbb{R}$ -Hilbert space and a map  $\phi : \mathcal{X} \rightarrow \mathcal{H}$  such that  $\forall x, x' \in \mathcal{X}$ ,

$$k(x, x') := \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}.$$

### Positive definite of an inner product in a Hilbert space

**Definition 2.6. (Positive definite function)** A symmetric function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is positive definite if  $\forall n \leq 1, \forall (a_1, \dots, a_n) \in \mathbb{R}^n$  and  $\forall (x_1, \dots, x_n) \in \mathbb{X}^n$ ,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j k(x_i, x_j) \geq 0.$$

The function  $k(., .)$  is strictly positive definite if for mutually distinct  $x_1, \dots, x_n \in \mathcal{X}$ , the equality only holds for  $a_1 = \dots = a_n = 0$ .

**Lemma 2.7.** Let  $\mathcal{H}$  be any Hilbert space,  $\mathcal{X}$  be a non-empty set and  $\phi : \mathcal{X} \rightarrow \mathcal{H}$ . Then  $k(x, y) := \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$  is a positive definite function.

### Reproducing kernel Hilbert space

**Definition 2.8. (Reproducing kernel Hilbert space).** Let  $\mathcal{H}$  be any Hilbert space of  $\mathbb{R}$ -valued functions defined on a non-empty set  $\mathcal{X}$ . A function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is called a reproducing kernel of  $\mathcal{H}$ , and  $\mathcal{H}$  is a RKHS, if  $k$  satisfies

1.  $\forall x \in \mathcal{X}, k(., x) \in \mathcal{H}$ ,
2.  $\forall x \in \mathcal{X}, \forall f \in \mathcal{H}, \langle f, k(., x) \rangle_{\mathcal{H}} = f(x)$ , (reproducing property).

**Theorem 2.4. (Moore-Aronszajn).** Every positive definite kernel  $k$  is associated with a unique RKHS  $\mathcal{H}$ .

## 2.3 Kernel on Wasserstein space $\mathcal{W}_2(\mathbb{R})$

### 2.3.1 The Wasserstein space on $\mathbb{R}$

Let us consider the set  $\mathcal{W}_2(\mathbb{R})$  of probability measures on  $\mathbb{R}$  with a finite moment of order two. For two  $\mu, \nu$  probability distributions in  $\mathcal{W}_2(\mathbb{R})$ , we denote  $\Pi(\mu, \nu)$  the

set of all probability measures  $\pi$  over the product set  $\mathbb{R} \times \mathbb{R}$  with first (resp. second) marginal  $\mu$  (resp.  $\nu$ ).

The transportation cost with quadratic cost function, which we denote quadratic transportation cost, between two measures  $\mu$  and  $\nu$  is defined as:

$$\mathcal{T}_2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int |x - y|^2 d\pi(x, y). \quad (2.3.1)$$

This transportation cost allows to endow the set  $\mathcal{W}_2(\mathbb{R})$  with a metric by defining the quadratic Monge-Kantorovich (or quadratic Wasserstein) distance between  $\mu$  and  $\nu$  as

$$W_2(\mu, \nu) = \mathcal{T}_2(\mu, \nu)^{1/2}.$$

A probability measure  $\pi$  in  $\Pi(\mu, \nu)$  performing the infimum in (2.3.1) is called an optimal coupling. This vocabulary transfers to a random vector  $(X_1, X_2)$  with distribution  $\pi$ . We will call  $\mathcal{W}_2(\mathbb{R})$  endowed with the distance  $W_2$  the Wasserstein space. More details on Wasserstein spaces and their links with optimal transport problems can be found in [Villani \(2008\)](#).

For distributions in  $\mathbb{R}$ , the Wasserstein distance can be written in a simpler way as follows: For any  $\mu \in \mathcal{W}_2(\mathbb{R})$ , we denote by  $F_\mu^{-1}$  the quantile function associated to  $\mu$ . Given a uniform random variable  $U$  on  $[0, 1]$ ,  $F_\mu^{-1}(U)$  is the random variable with law  $\mu$ . Then, for every  $\mu$  and  $\nu$  the random vector  $(F_\mu^{-1}(U), F_\nu^{-1}(U))$  is the optimal coupling (see [Villani \(2008\)](#)), where  $F^{-1}$  is defined as

$$F_\mu^{-1}(t) = \inf\{u, F_\mu(u) \geq t\}. \quad (2.3.2)$$

In this case, the simplest expression for the Wasserstein distance is given in [Whitt \(1976\)](#):

$$W_2(\mu, \nu) = \mathbb{E}(F_\mu^{-1}(U) - F_\nu^{-1}(U))^2. \quad (2.3.3)$$

Topological properties of Wasserstein spaces are reviewed in [Villani \(2008\)](#). Hereafter, compactity will be required and will be obtained as follows: let  $\Omega \subset \mathbb{R}$  be a compact subset, then the Wasserstein space  $\mathcal{W}_2(\Omega)$  is also compact. In this paper, we consider Wasserstein spaces  $\mathcal{W}_2(\Omega)$ , where  $\Omega$  is a compact subset on  $\mathbb{R}$  endowed with the Wasserstein distance  $W_2$ . Hence for any  $\mu \in \mathcal{W}_2(\Omega)$ , we denote  $F_\mu|_\Omega : \Omega \rightarrow [a, b]$  with  $[a, b] \subset [0, 1]$  the distribution function restricted on a compact subset  $\Omega$ . We also define  $F_\mu^{-1}|_\Omega$  as:

$$F_\mu^{-1}|_\Omega(t) = \inf\{u \in \Omega, F_\mu|_\Omega(u) \geq t\}, \forall t \in [a, b]. \quad (2.3.4)$$

Given a uniform random variable  $V$  on  $[a, b]$ ,  $F_\mu^{-1}|_\Omega$  is a random variable with law  $\mu$ . By inheriting properties from  $\mathcal{W}_2(\mathbb{R})$  for every  $\mu$  and  $\nu$  in  $\mathcal{W}_2(\Omega)$ , the random vector  $(F_\mu^{-1}|_\Omega(V), F_\nu^{-1}|_\Omega(V))$  is an optimal coupling. In this case, we consider in this paper the simplest expression for the Wasserstein distance between  $\mu$  and  $\nu$  in  $\mathcal{W}_2(\Omega)$ :

$$W_2(\mu, \nu) = \mathbb{E}(F_\mu^{-1}|_\Omega(V) - F_\nu^{-1}|_\Omega(V))^2. \quad (2.3.5)$$

### 2.3.2 Kernel

Constructing a positive definite kernel defined on the Wasserstein space is not obvious and was recently done in [Bachoc et al. \(2017\)](#). For sake of completeness, we recall here briefly this construction.

**Theorem 2.5.** *Let  $k_\Theta : \mathcal{W}_2(\Omega) \times \mathcal{W}_2(\Omega) \rightarrow \mathbb{R}$  with the parameter  $\Theta := (\gamma, H, l)$  such that  $\gamma \neq 0$  and  $l > 0$  defined as*

$$k_\Theta(\mu, \nu) := \gamma^2 \exp\left(-\frac{W_2^{2H}(\mu, \nu)}{l}\right). \quad (2.3.6)$$

*Then for  $0 < H \leq 1$ ,  $k_\Theta$  is a positive definite kernel.*

The proof of this Theorem directly follows Theorem 2.6 and Propositions 2.7. In this paper we use Theorem 2.5 to study the properties of such kernel in the RKHS regression framework.

**Theorem 2.6. (Schoenberg)** *Let  $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  be a completely monotone function, and  $K$  a negative definite kernel. Then  $(x, y) \mapsto F(K(x, y))$  is a positive definite kernel.*

*Proof.* We refer to [Cowling \(1983\)](#) for more details on completely monotone functions and a proof of Theorem 2.6. We use this result to give a proof of the positive definite kernel which is presented in Theorem 2.5.  $\square$

The following proposition which can be found a general version for any distribution in  $\mathcal{W}_2(\mathbb{R})$  in [Bachoc et al. \(2017\)](#), here we present again a result for our case in  $\mathcal{W}_2(\Omega)$ . Finally we give conditions on the exponent  $H$  to achieve the negative definite kernel using exponents of the Wasserstein distance.

**Proposition 2.7.** *The function  $W_2^{2H}$  is a negative definite kernel if and only if  $0 < H \leq 1$ .*

*Proof.* One can find in [Kolouri et al. \(2015\)](#) a version of this result for absolutely continuous distributions in  $\mathcal{W}_2(\Omega)$ . The proof of this proposition is proved for any distribution in  $\mathcal{W}_2(\Omega)$ .

For any  $\mu \in \mathcal{W}_2(\Omega)$ , we denote by  $F_\mu^{-1}|_\Omega$  the function defined by (2.3.4).

From statements in Section 2.3.1, we well known that given a uniform random variable  $V$  on  $[a, b]$ ,  $F_\mu^{-1}|_\Omega(V)$  is a random variable with a law  $\mu$ , and furthermore for every  $\mu$  and  $\nu$  in  $\mathcal{W}_2(\Omega)$ ,

$$W_2(\mu, \nu) = \mathbb{E}(F_\mu^{-1}|_\Omega(V) - F_\nu^{-1}|_\Omega(V))^2,$$

that is to say the coupling of  $\mu$  and  $\nu$  given by the random vector  $(F_\mu^{-1}|_\Omega(V), F_\nu^{-1}|_\Omega(V))$  is optimal.

Now let us consider every  $\mu_1, \dots, \mu_n \in \mathcal{W}_2(\Omega)$  and  $c_1, \dots, c_n \in \mathbb{R}$  such that  $\sum_{i=1}^n c_i = 0$ . We have

$$\begin{aligned} \sum_{i=1}^n c_i c_j W_2^2(\mu_i, \mu_j) &= \sum_{i,j=1}^n c_i c_j \mathbb{E}(F_{\mu_i}^{-1}|_{\Omega}(V) - F_{\mu_j}^{-1}|_{\Omega}(V))^2 \\ &= \sum_{i,j=1}^n c_i c_j \mathbb{E}\left(F_{\mu_i}^{-1}|_{\Omega}(V)\right)^2 + \sum_{i,j=1}^n c_i c_j \mathbb{E}\left(F_{\mu_j}^{-1}|_{\Omega}(V)\right)^2 \\ &\quad - 2 \sum_{i,j=1}^n c_i c_j \mathbb{E}\left(F_{\mu_i}^{-1}|_{\Omega}(V) F_{\mu_j}^{-1}|_{\Omega}(V)\right). \end{aligned}$$

Using  $\sum_{i=1}^n c_i = 0$  the first two sums vanish and we obtain

$$\begin{aligned} \sum_{i=1}^n c_i c_j W_2^2(\mu_i, \mu_j) &= -2 \sum_{i,j=1}^n c_i c_j \mathbb{E}\left(F_{\mu_i}^{-1}|_{\Omega}(V) F_{\mu_j}^{-1}|_{\Omega}(V)\right) \\ &= -2 \mathbb{E}\left(\sum_{i=1}^n c_i F_{\mu_i}^{-1}|_{\Omega}(V)\right)^2 \leq 0. \end{aligned}$$

Thus  $W_2^{2H}$  is a negative definite kernel for  $0 < H \leq 1$ .  $\square$

Now, we give the proof for Theorem 2.5 by using results from both of Schoenberg theorem and proposition 2.7.

*Proof.* The proof of Theorem 2.5 follows immediately below from Theorem 2.6 and Proposition 2.7. Applying Proposition 2.7, we deduce that  $W_2^{2H}(\mu, \nu)$  with  $H = 1$  is a negative definite kernel for all  $\mu, \nu$  in  $\mathcal{W}_2(\Omega)$ .

We can easily see that  $e^{-\lambda x}$  with  $\lambda$  positive is a completely monotone function. Let us then consider a mapping as follows:

$$\begin{aligned} F : \mathbb{R}^+ &\rightarrow \mathbb{R}^+ \\ x &\mapsto \gamma^2 e^{-\lambda x}, \end{aligned}$$

where  $\gamma^2 > 0$ ,  $x = W_2^2(\mu, \nu)$  with  $\lambda = \frac{1}{l}$ ,  $l > 0$ . Then  $F$  is also a completely monotone function. From Theorem 2.6,  $k_{\Theta}$  is a positive definite kernel.  $\square$

## 2.4 Regression

### 2.4.1 Setting

In this section, we aim to define a regression function with distribution inputs. The problem of distribution regression consists in estimating an unknown function  $f : \mathcal{W}_2(\Omega) \rightarrow \mathbb{R}$  by using observations  $(\mu_i, y_i)$  in  $\mathcal{W}_2(\Omega) \times \mathbb{R}$  for all  $i = 1, \dots, n$ . We recall observes in (2.1.1) as follows

$$y_i = f(\mu_i) + \epsilon_i. \quad (2.4.1)$$

To provide a general form for functions defined on distributions, we will use the RKHS framework. Let  $k_\Theta : \mathcal{W}_2(\Omega) \times \mathcal{W}_2(\Omega) \mapsto \mathbb{R}$  be defined in Theorem 2.5. For a fixed valid  $\Theta$ , we define a space  $\mathcal{F}_0$  as follows:

$$\mathcal{F}_0 := \text{span} \{k_\Theta(\bullet, \mu) : \mu \in \mathcal{W}_2(\Omega)\}.$$

And  $\mathcal{F}_0$  is endowed with the inner product

$$\langle f_n, g_m \rangle_{\mathcal{F}_0} = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j k_\Theta(\mu_i, \nu_j),$$

where  $f_n(\bullet) = \sum_{i=1}^n \alpha_i k_\Theta(\bullet, \mu_i)$  and  $g_m(\bullet) = \sum_{j=1}^m \beta_j k_\Theta(\bullet, \nu_j)$ . The norm in  $\mathcal{F}_0$  corresponds to the inner product,

$$\|f_n\|_{\mathcal{F}_0}^2 = \sum_{i=1}^n \alpha_i^2 k_\Theta(\mu_i, \mu_i). \quad (2.4.2)$$

Let  $\mathcal{F}$  be the space of all continuous real-valued functions from  $\mathcal{W}_2(\Omega)$  to  $\mathbb{R}$  while  $\mathcal{F}_0$  consists of all functions in  $\mathcal{F}$  which are uniform limits of functions of form  $f_n$ . The kernel  $k_\Theta$  is said to be universal if it has a property that  $\mathcal{F}_0 = \mathcal{F}$ .

From that for all  $f, g$  belong in  $\mathcal{F}$ , the inner product is well defined as following formula

$$\langle f, g \rangle_{\mathcal{F}} := \lim_{n \rightarrow \infty} \langle f_n, g_n \rangle_{\mathcal{F}_0}. \quad (2.4.3)$$

Coming back to our problem, we want to estimate the unknown function  $f$  by an estimation function  $\hat{f}$  obtained by minimizing the regularized empirical risk over the RKHS  $\mathcal{F}$ . For this consider, we solve the solution of the minimization problem

$$\hat{f} = \underset{f \in \mathcal{F}}{\text{argmin}} \left( \sum_{i=1}^n |y_i - f(\mu_i)|^2 + \lambda \|f\|_{\mathcal{F}}^2 \right), \quad (2.4.4)$$

where  $\lambda \in \mathbb{R}^+$  is the regularization parameter. Using the Representer theorem, this leads to the following expression for  $\hat{f}$ ,

$$\hat{f} : \mu \mapsto \hat{f}(\mu) := \sum_{j=1}^n \hat{\alpha}_j k_\Theta(\mu, \mu_j), \quad (2.4.5)$$

where  $\{\hat{\alpha}_j\}_{j=1}^n$  are parameters typically obtained from training data.

## 2.4.2 Universal kernel

First, we recall the definition of a universal kernel and the main theorem to ensure universal properties of positive definite kernels in Theorem 2.5.

**Definition 2.9.** *Let  $C(X)$  be the space of continuous bounded functions on compact domain  $X$ . A continuous kernel  $k$  on domain  $X$  is called universal if the space of all functions induced by  $k$  is dense in  $C(X)$ , i.e, for all  $f \in C(X)$  and every  $\epsilon > 0$  there exists a function  $g$  induced by  $k$  with  $\|f - g\|_\infty \leq \epsilon$ .*

For more information on universal kernel and RKHS, we refer to Chapter 4 in [Villani \(2008\)](#) and [Sriperumbudur et al. \(2011\)](#), [Micchelli, Xu, and Zhang \(2006\)](#).

Now, we introduce a main contribution of our work which shows that kernels based on Wassertein distance proposed in [Bachoc et al. \(2017\)](#) are universal, finally, we get what kernel we need for the setting [2.4.1](#).

**Theorem 2.8.** *Let choose the parameter  $\Theta$  in [\(2.3.6\)](#) such that  $\gamma \neq 0$ ,  $l > 0$  and  $H = 1$ . The kernel  $k_\Theta : \mathcal{W}_2(\Omega) \times \mathcal{W}_2(\Omega) \rightarrow \mathbb{R}$  defined in [\(2.3.6\)](#) is universal.*

To prove this Theorem, we need the two following Proposition [2.9](#) and Proposition [2.10](#). We will give the proof of this result after the two of following statements.

**Proposition 2.9.** *Let an increasing function  $F_\mu|_\Omega : \Omega \rightarrow [a, b]$  with  $[a, b] \subset [0, 1]$  be the distribution function restricted on a compact subset  $\Omega$  of  $\mathbb{R}$ ,  $F_\mu^{-1}|_\Omega$  be defined by  $F_\mu^{-1}|_\Omega(t) = \inf\{u \in \Omega, F_\mu|_\Omega(u) \geq t\}$  for all  $t \in [a, b]$ . Then  $F_\mu|_\Omega$  is continuous if and only if  $F_\mu^{-1}|_\Omega$  is strictly increasing on  $[\inf \text{ran} F_\mu|_\Omega, \sup \text{ran} F_\mu|_\Omega]$ .  $F_\mu|_\Omega$  is strictly increasing if and only if  $F_\mu^{-1}|_\Omega$  is continuous on  $\text{ran} F_\mu|_\Omega$ , where  $\text{ran} F_\mu|_\Omega := \{F_\mu|_\Omega(x) : x \in \Omega\}$  denotes the range of  $F_\mu|_\Omega$ .*

*Proof.* For the proof of this proposition, we refer to [Embrechts and Hofert \(2013\)](#) for more theoretical details in the case of the quantile function from  $[0, 1]$  to  $\mathbb{R}$ . Here, we give the proof for the kind of function  $F_\mu^{-1}|_\Omega$  from  $[a, b] \subset [0, 1]$  to  $\Omega$  be a compact subset of  $\mathbb{R}$ .

First, we prove that  $F_\mu|_\Omega$  is continuous  $\Leftrightarrow F_\mu^{-1}|_\Omega$  is strictly increasing on  $[\inf \text{ran} F_\mu|_\Omega, \sup \text{ran} F_\mu|_\Omega]$ .

$\Rightarrow$ ) Assume that  $F_\mu|_\Omega$  is discontinuous at  $x_0 \in \Omega$ , we need to prove that  $F_\mu^{-1}|_\Omega$  is not strictly increasing on  $[\inf \text{ran} F_\mu|_\Omega, \sup \text{ran} F_\mu|_\Omega]$ .

In fact, since  $F_\mu|_\Omega$  is increasing, this implies that there exist

$$\begin{aligned} z_1 &:= F_\mu|_\Omega(x_0-) := \lim_{x \uparrow x_0} F_\mu|_\Omega(x) \\ z_2 &:= F_\mu|_\Omega(x_0+) := \lim_{x \downarrow x_0} F_\mu|_\Omega(x), \end{aligned}$$

$z_1 < z_2$  and  $z_0 := F_\mu|_\Omega(x_0) \in [z_1, z_2]$  such that  $z_1 < z_0 < z_2$ .

Now, there exists  $z_3, z_4 \in [a, b]$  such that either  $z_1 < z_3 < z_4 < z_0 < z_2$  or  $z_1 < z_0 < z_3 < z_4 < z_2$  and note that for all  $z \in [z_3, z_4] \in [\inf \text{ran} F_\mu|_\Omega, \sup \text{ran} F_\mu|_\Omega]$ ,  $z \notin \text{ran} F_\mu|_\Omega$ .

From that, by definition of  $F_\mu^{-1}|_\Omega$ , it is a constant in  $[z_3, z_4]$ . Hence,  $F_\mu^{-1}|_\Omega$  is not strictly increasing on  $[\inf \text{ran} F_\mu|_\Omega, \sup \text{ran} F_\mu|_\Omega]$ .

$\Rightarrow$ ) Assume that  $F_\mu^{-1}|_\Omega$  be not strictly increasing on  $[\inf \text{ran} F_\mu|_\Omega, \sup \text{ran} F_\mu|_\Omega]$ , we need to prove that  $F_\mu|_\Omega$  is discontinuous.

In fact, because  $F_\mu^{-1}|_\Omega$  is not strictly increasing on  $[\inf \text{ran} F_\mu|_\Omega, \sup \text{ran} F_\mu|_\Omega]$ , there exists  $[z_1, z_2] \in \Omega$  with  $\inf \text{ran} F_\mu|_\Omega < z_1 < z_2 < \sup \text{ran} F_\mu|_\Omega$  such that  $F_\mu^{-1}|_\Omega(z) = x$ ,  $\forall z \in [z_1, z_2]$  and a  $x \in \Omega$ .

By definition of  $F_\mu^{-1}|_\Omega$ , this implies that

$$F_\mu|_\Omega(x - \epsilon) < z_1 < z_2 < F_\mu|_\Omega(x + \epsilon),$$



for all  $\epsilon > 0$ .

Letting  $\epsilon \downarrow 0$ , we obtain

$$F_\mu|_\Omega(x-) \leq z_1 < z_2 \leq F_\mu|_\Omega(x+).$$

Thus  $F_\mu|_\Omega$  is discontinuous at  $x$ .

Now, we prove the second statement:  $F_\mu|_\Omega$  is not strictly increasing  $\Leftrightarrow F_\mu^{-1}|_\Omega$  is discontinuous on  $\text{ran}F_\mu|_\Omega$ .

To prove this statement, we need to use a result (7) in Proposition 2.3 in [Embrechts and Hofert \(2013\)](#), we note this result by (\*). Here, we recall briefly this result corresponding to our case as follows.

$$\left(F_\mu^{-1}|_\Omega(z-), F_\mu^{-1}|_\Omega(z+)\right) \subseteq \{x \in \Omega : F_\mu|_\Omega(x) = z\} \subseteq \left[F_\mu^{-1}|_\Omega(z-), F_\mu^{-1}|_\Omega(z+)\right],$$

where

$$\begin{aligned} F_\mu^{-1}|_\Omega(z-) &= \lim_{t \uparrow z} F_\mu^{-1}|_\Omega(t), \\ F_\mu^{-1}|_\Omega(z+) &= \lim_{t \downarrow z} F_\mu^{-1}|_\Omega(t). \end{aligned}$$

$\Rightarrow$ ) Now, assume that  $F_\mu|_\Omega$  is not strictly increasing, we prove that  $F_\mu^{-1}|_\Omega$  is discontinuous on  $\text{ran}F_\mu|_\Omega$ .

In fact, let  $A_z := \{x \in \Omega : F_\mu|_\Omega(x) = z\}$ .

Because  $F_\mu|_\Omega$  is not strictly increasing, there exists a  $x \in \Omega$  such that  $A_z$  contains an open interval.

From the second statement of above result (\*), we have

$$F_\mu^{-1}|_\Omega(z-) < F_\mu^{-1}|_\Omega(z+).$$

Thus  $F_\mu^{-1}|_\Omega$  is discontinuous at  $z \in \text{ran}F_\mu|_\Omega$ .

$\Leftarrow$ ) Assume  $F_\mu^{-1}|_\Omega$  is discontinuous on  $\text{ran}F_\mu|_\Omega$ , we prove that  $F_\mu|_\Omega$  is not strictly increasing.

In fact, because  $F_\mu^{-1}|_\Omega$  is discontinuous on  $\text{ran}F_\mu|_\Omega$ , there exists  $z \in \text{ran}F_\mu|_\Omega$  such that  $F_\mu^{-1}|_\Omega(z-) < F_\mu^{-1}|_\Omega(z+)$ . Using the second statement in the result (\*) that  $A_z$  contains the open interval  $\left(F_\mu^{-1}|_\Omega(z-), F_\mu^{-1}|_\Omega(z+)\right)$  thus  $F_\mu|_\Omega$  is not strictly increasing.  $\square$

**Proposition 2.10.** *Let  $X$  be a compact metric space and  $\mathcal{H}$  be a separable Hilbert space such that there exist a continuous and injective map:  $\rho : X \rightarrow \mathcal{H}$ . For  $\gamma > 0$ , the Gaussian-type RBF-kernel  $k_\gamma : X \times X \rightarrow \mathbb{R}$  is the universal kernel, where*

$$k_\sigma(x, x') := \exp(-\sigma^2 \|\rho(x) - \rho(x')\|_{\mathcal{H}}^2), \quad x, x' \in X.$$

*Proof.* See a part iii) of Theorem 2.2 in [Christmann and Steinwart \(2010\)](#) for the proof.  $\square$

Now, using results of the two above results from Proposition 2.9 and 2.10, we give a proof for Theorem 2.8.

*Proof.* From Proposition 2.9 with the conditions including the distribution restricted on  $\Omega$ ,  $F_\mu|_\Omega$  be continuous and  $F_\mu|_\Omega$  be strictly increasing on  $\Omega$ , then there exists a continuous and injective map

$$\begin{aligned}\rho : \mathcal{W}_2(\Omega) &\rightarrow \mathbb{L}_2[a, b] \\ \mu &\mapsto \rho(\mu) := F_\mu^{-1}|_\Omega.\end{aligned}$$

$F_\mu^{-1}|_\Omega$  is continuous on  $[a, b]$  and strictly increasing on  $[\inf \text{ran} F_\mu|_\Omega, \sup \text{ran} F_\mu|_\Omega]$ . We consider a Wasserstein space  $\mathcal{W}_2(\Omega)$  metrized by the Wasserstein distance  $W_2$  with  $\Omega$  be a compact subset on  $\mathbb{R}$  and  $\mathbb{L}_2[a, b]$  be the usual space of square integrable functions on  $[a, b]$ . For  $\sigma$  in Proposition 2.10 is exactly defined by  $1/\sqrt{l}$  for all  $l > 0$ . We have

$$k_\Theta(\mu, \nu) = \gamma^2 \exp \left\{ -\frac{\|F_\mu^{-1}|_\Omega - F_\nu^{-1}|_\Omega\|_{\mathbb{L}_2[a, b]}^2}{l} \right\}$$

is the universal kernel from Proposition 2.10. We complete the proof of Theorem 2.8.  $\square$

The minimization program in (2.4.4) can be solved explicitly using the representer theorem of Kimeldorf and Wahba (1971). Note that Schölkopf and Smola Smola and Schölkopf (1998) give a simple proof of a more general version of the theorem. Define  $c_{ij}$  as follows

$$c_{ij} = \gamma^2 \exp \left( -\frac{W_2^2(\mu_i, \mu_j)}{l} \right)$$

and  $\alpha = (\alpha_1, \dots, \alpha_n)^T$ ,  $Y = (y_1, \dots, y_n)^T$ .

Now we take the matrix formulation of (2.4.4) we obtain

$$\min_{\alpha} \text{trace}((Y - C\alpha)(Y - C\alpha)^T) + \lambda \text{trace}(C\alpha\alpha^T), \quad (2.4.6)$$

where the operation trace is defined as

$$\text{trace}(A) = \sum_{i=1}^n a_{ii}$$

with  $A = (a_{ii})_{i=1}^n$ .

Taking the derivative of (2.4.6) with respect to vector  $\alpha$ , we find that  $\alpha$  satisfies the system of linear equations

$$(C + \lambda I)\alpha = Y. \quad (2.4.7)$$

Hence

$$\hat{f}(\mu) = \sum_{j=1}^n \hat{\alpha}_j k_\Theta(\mu, \mu_j), \quad (2.4.8)$$

with

$$\hat{\alpha} = (C + \lambda I)^{-1}Y. \quad (2.4.9)$$

## 2.5 Numerical Simulations and Real data application

First we consider the regression model with one-dimensional distribution input. For this, we train the model and try to estimate the unknown function  $f$  using our proposed methodology. Then we address the regression model but with higher-dimensional distributions entries. By using the same technique for generating data in the one-dimensional distribution, we estimate the unknown function  $f$  in the case of 2-dimensional distributions. Finally, we apply our for a real dataset from biological field.

### 2.5.1 Simulation in the regression model with one-dimensional distribution input

#### Overview of the simulation procedure

In this section, we investigate the regression model for predicting the regression function from distributions. Particularly, we want to estimate the unknown function  $f$  in model (2.4.1) by using the proposed estimation  $\hat{f}$  in (2.4.8), so we need to present how we can optimize the parameters in this formula. We then compare the regression model based on RKHS induced by our universal kernel function to more classical kernel functions operating on projections of the probability measures on finite dimensional spaces. We address the input-output map given by

$$f(\nu) = \frac{m_\nu}{0.05 + \sigma_\nu}, \quad (2.5.1)$$

where  $\nu$  is a Gaussian distribution of mean  $m_\nu$  and variance  $\sigma_\nu^2$ . We consider the ground truth function  $f$  that we compare with a predicted function  $\hat{f}$ , such as:

$$\hat{f}(\nu) = \gamma^2 \sum_{j=1}^n \hat{\alpha}_j \exp \left[ -\frac{W_2^2(\nu, \mu_j)}{l} \right], \quad (2.5.2)$$

where the Wasserstein distance between two Gaussian distribution is calculated using:

$$W_2^2(\mu, \nu) = (m_\mu - m_\nu)^2 + (\sigma_\mu - \sigma_\nu)^2,$$

where  $\mu \sim \mathcal{N}(m_\mu, \sigma_\mu^2)$  and  $\nu \sim \mathcal{N}(m_\nu, \sigma_\nu^2)$ .

Each value  $\hat{\alpha}_j$  is estimated using Eq. (2.4.9) which depends on parameter  $\lambda > 0$ . Thus our proposed estimation function  $\hat{f}$  depends totally on the three parameters  $\lambda > 0, \gamma \neq 0$  and  $l > 0$ . To understand the effects of these parameters on  $\hat{f}$ , we define reference values of  $\lambda, \gamma$  and  $l$ . We then generate a training set including the normal distributions  $\nu_i \sim \mathcal{N}(m_{\nu_i}, \sigma_{\nu_i})$  such that  $cor(\nu_i, \nu_j) \neq 0, \forall i, j = 1, \dots, n$ , with  $n$  be a size of training set. In this simulation, we take  $n = 200$ .

From the training set  $\{(\nu_i, f(\nu_i))\}_{i=1}^{n=200}$ , we fit three regression models which we

call "Wasserstein", "Legendre" and "Histogram". "Wasserstein" stands for our model, "Legendre" corresponds to regression models with smooth density input obtained by projecting onto Legendre polynomial basis while the last "Histogram" is introduced for the rough histogram input, as detailed later. Then we evaluate the quality of the two regression models on a test set of size  $n_t$  of the form  $\{(\nu_{t,i}, f(\nu_{t,i}))\}_i^{n_t}$ , where  $\nu_{t,i}$  is generated in the same way as  $\nu_i$  above. We consider the following quality criteria, that is the root mean square error (RMSE) to see the quality of our regression model by

$$RMSE^2(\hat{f}, f) = \frac{1}{n_t} \sum_{i=1}^{n_t} [f(\nu_{t,i}) - \hat{f}(\nu_{t,i})]^2.$$

### Detail on the regression models

We refer to our model as "Wasserstein" and introduce briefly "Legendre" and "Histogram" regression models.

The "Wasserstein" model first propose the estimated function as follows

$$\hat{f}(\nu_{t,i}) = \gamma^2 \sum_{j=1}^n \hat{\alpha}_j \exp \left[ -\frac{(m_{\nu_j} - m_{\nu_{t,i}})^2 + (\sigma_{\nu_j} - \sigma_{\nu_{t,i}})^2}{l} \right], \quad (2.5.3)$$

where  $\nu_{t,i}, i = 1, \dots, n_t$  belong to testing set with size  $n_t$ , and  $\nu_j, j = 1, \dots, n$  belong to training set with size  $n$ . The estimated function  $\hat{f}$  depends on three parameters  $\gamma \neq 0$ ,  $\lambda > 0$  and  $l > 0$ .

The "Legendre" model is based on kernel functions operating on finite dimensional linear projections of the distributions. For a Gaussian distribution  $\mu \sim \mathcal{N}(m, \sigma^2)$  with density  $f_\mu(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right)$  and support  $[0, 1]$ , we compute for  $i = 0, \dots, \theta - 1$ :

$$a_i(\mu) = \int_0^1 \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) p_i(t) dt,$$

where  $p_i$  is the  $i$ -th normalized Legendre polynomial, with  $\int_0^1 p_i^2(t) dt = 1$ . The integer  $\theta$  is called the order of the decomposition. Then  $k_L$  operators on the vector  $(a_0(\nu), \dots, a_{\theta-1}(\nu))$  and is of the form

$$k_L(\nu_1, \nu_2) = \gamma^2 \exp \left[ -\sum_{i=0}^{\theta-1} \frac{|a_i(\nu_1) - a_i(\nu_2)|}{l_i} \right]. \quad (2.5.4)$$

Thus the estimated regression function  $\hat{f}$  in this case is calculated by following function

$$\hat{f}(\nu_{t,i}) = \gamma^2 \sum_{j=1}^n \hat{\alpha}_j \exp \left[ -\sum_{i=0}^{\theta-1} \frac{|a_i(\nu_{t,i}) - a_i(\nu_j)|}{l_i} \right].$$

We just consider the orders of the decomposition 5 and 10. We fix  $l_i = l$  for all  $i = 1, \dots, \theta - 1$ , this estimated function depends also on three parameters  $\gamma \neq 0$ ,  $\lambda > 0$  and  $l > 0$ .

The "Histogram" model is based on a Gaussian RBF kernel with  $\chi_2$ -distance between the histograms, see [Vedaldi, Gulshan, Varma, and Zisserman \(2009\)](#). More particularly, for  $0 < \zeta \leq 1$ ,

$$K(\nu_1, \nu_2) = \exp(-\zeta \chi_2(h_1, h_2)),$$

where

$$\chi_2(h_1, h_2) = \sum_{r=1}^M \frac{(h_1(r) - h_2(r))^2}{h_1(r) + h_2(r)},$$

with  $h_i$  is the histogram of the  $i$ th Gaussian distribution  $\nu_i$ ,  $i = \{1, 2\}$  and  $M$  is the number of bins for all histograms  $h_i$ .

Thus, in this case, the estimated regression function is considered by the following expression

$$\hat{f}(\nu_{i,i}) = \sum_{j=1}^n \hat{\alpha}_j \exp\left(-\zeta \sum_{r=1}^M \frac{(h_1(r) - h_2(r))^2}{h_1(r) + h_2(r)}\right).$$

## Result

In the simulations, we observe the effects of parameters  $\lambda > 0$ ,  $\gamma \neq 0$  and  $l > 0$  on RMSE between predicted function  $\hat{f}(\nu_{i,t})$  and exact function  $f(\nu_{i,t})$  through the testing set  $\{(\nu_{i,t})\}_{i=1}^{n_t}$ . We also take two sizes of testing set  $n_t = 500, n_t = 700$  to see the changes of RMSE. We just show the detailed presentation about choosing the optimal parameters on the "Wasserstein" model.

**Case of testing set size  $n_t = 500$**  : Now we consider RMSE in the case of  $n_t = 500$  under the different fixed values  $\gamma = 1/2, 1, 10$  and running  $\lambda > 0$  separated by 30 values from 0.005 to 30,  $l > 0$  separated by 25 values from 0.005 to 20. Let us see here the values of RMSE with the different cases of  $\gamma$  in following [Figure 2.1, 2.2, 2.3](#).

We realize through three choices of  $\gamma = 1/2, 1, 10$  that the values  $\gamma$  give the same impact of RMSE variations, but the smallest RMSE in the case  $\gamma = 1$ . In following this stimulation, we fix the value of  $\gamma = 1$  and run the values of  $\lambda$  and  $l$  to see the changes of RMSE in the case of bigger size of testing set.

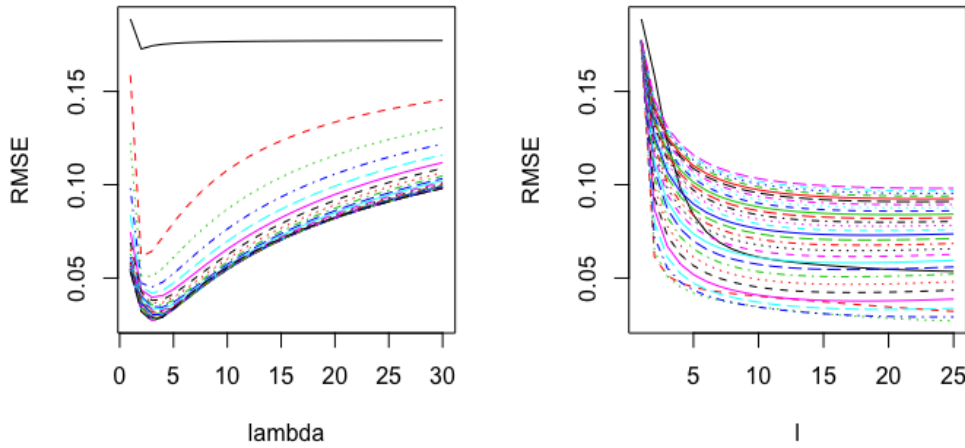


Figure 2.1: In the case of  $n_t = 500$ , fixing a value  $\gamma = 1/2$ , we run  $\lambda > 0$  separated by 30 values from 0.005 to 30,  $l > 0$  separated by 25 values from 0.005 to 20. We see the two graphs, one follows values of  $\lambda$  in the left side and  $l$  in the right side. RMSE will be minimized, in which it is lower than 0.08, with  $0 < \lambda < 15$  and  $l$  big enough. We note that when  $0 < l < 1$  RMSE is quite a big value for all  $\lambda > 0$ , so we avoid to chose these values of  $l$ .

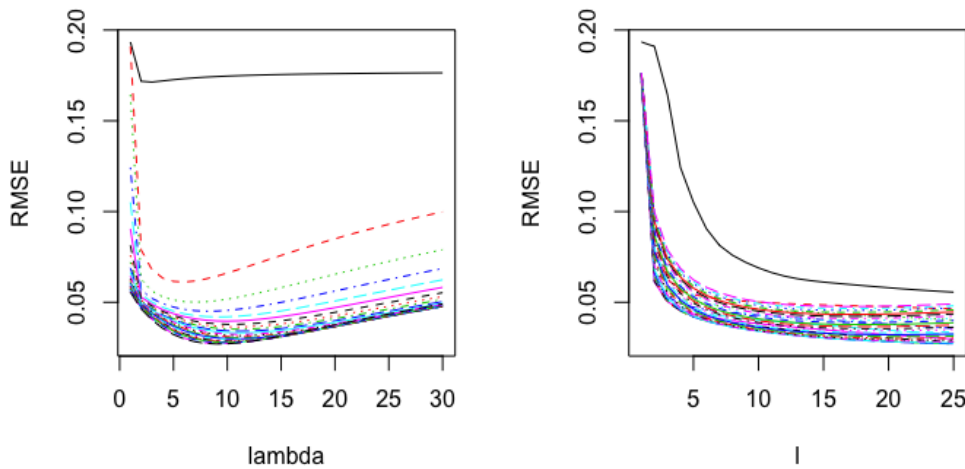


Figure 2.2: In the case of  $n_t = 500$ , fixing a value  $\gamma = 1$ , we run  $\lambda > 0$  separated by 30 values from 0.005 to 30,  $l > 0$  separated by 25 values from 0.005 to 20. We see the two graphs, one follows values of  $\lambda$  in the left side and  $l$  in the right side. The variations of RMSE in this case is not change significantly with the case of  $\gamma = 1/2$ , however, it looks smaller than the case  $\gamma = 1/2$ . We also see that RMSE will be minimized by two case: first  $0 < \lambda < 1$  and  $l$  big enough; second  $\lambda > 1$  for all  $l > 1$  and RMSE is quite big at  $0 < l < 1$  for all  $\lambda > 0$ .

**Case of testing set size  $n_t = 700$  :** Now we consider RMSE in the case of  $n_t = 700$  in Figure 2.4 for a fixed value  $\gamma = 1$  and running  $\lambda > 0$  separated by 30 values from 0.005 to 30,  $l > 0$  separated by 25 values from 0.005 to 20. The parameters  $\lambda$  and  $l$  control the smoothness of the estimator. See the Figure 2.5, to more clearly about our regression model with the exact function defined in (2.5.1).

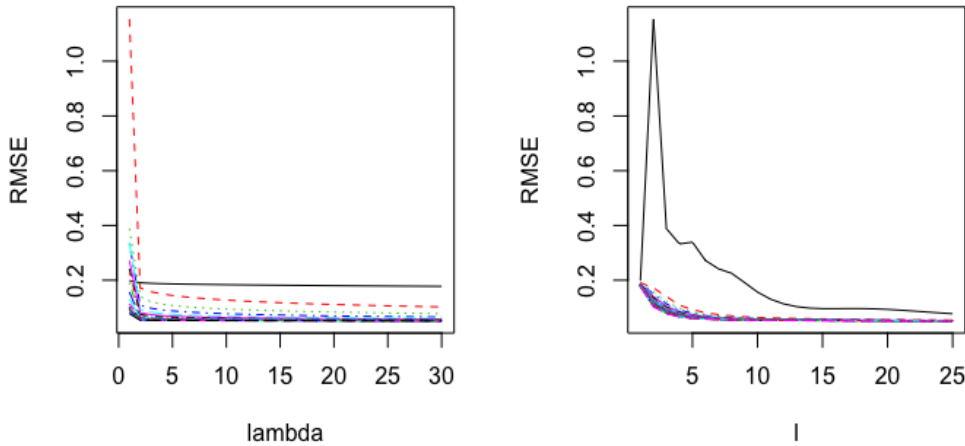


Figure 2.3: In the case of  $n_t = 500$ , fixing a value  $\gamma = 10$ , we run  $\lambda > 0$  separated by 30 values from 0.005 to 30,  $l > 0$  separated by 25 values from 0.005 to 20. RMSE in this case looks bigger than two above cases of  $\gamma = 1/2$  and  $\gamma = 1$ .

And finally, we consider the different RMSE's between "Wasserstein" and "Legendre" model by choosing the optimal values  $\gamma = 1$ ,  $\lambda = 5$  and  $l = 10$  under considering  $n_t = 700$ . In Table 3.1, we show the values of RMSE quality criteria for the "Wasserstein" and "Legendre" distribution regression models. From the values of the RMSE criterion, the "Wasserstein" model clearly outperforms the other models. The RMSE of the "Legendre" models slightly decreases when the order increases, and stay well above the RMSE of the "Wasserstein" model.

model	RMSE
"Wasserstein"	0.04
"Legendre" order 5	0.15
"Legendre" order 10	0.11
"Histogram"	0.12

Table 2.1: RMSE values of the quality criteria for the "Wasserstein", "Legendre" and "Histogram" distribution regression models. The "Wasserstein" is based on universal kernel function operating directly on the input Gaussian distributions, while "Legendre" is based on linear projections of the Gaussian distribution inputs on finite-dimensional spaces. For "Legendre", the order value is the dimension of the projection space. Beside that, the "Histogram" is produced by the rough histograms of Gaussian distributions. The quality criteria is the root mean square error (RMSE) should be minimal. The "Wasserstein" distribution regression model clearly outperforms the "Legendre" with two orders 5, 10 and the "Histogram".

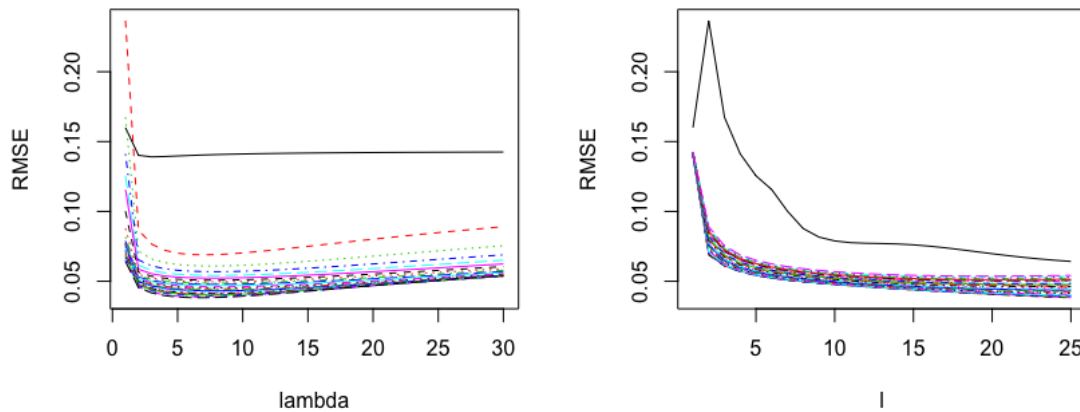


Figure 2.4: In the case of  $n_t = 700$ , fixing a value  $\gamma = 1$ , we run  $\lambda > 0$  separated by 30 values from 0.005 to 30,  $l > 0$  separated by 25 values from 0.005 to 20. RMSE is almost lower than 0.06 when  $\lambda > 1$  and  $l > 1$ . However for  $0 < \lambda < 1$ , we can also obtain the small RMSE when  $l$  big enough. This figure provides a view about size of testing set, in which for the big enough of testing set size we will obtain the smaller RMSE under of the optimal parameters  $\gamma$ ,  $\lambda$  and  $l$ .

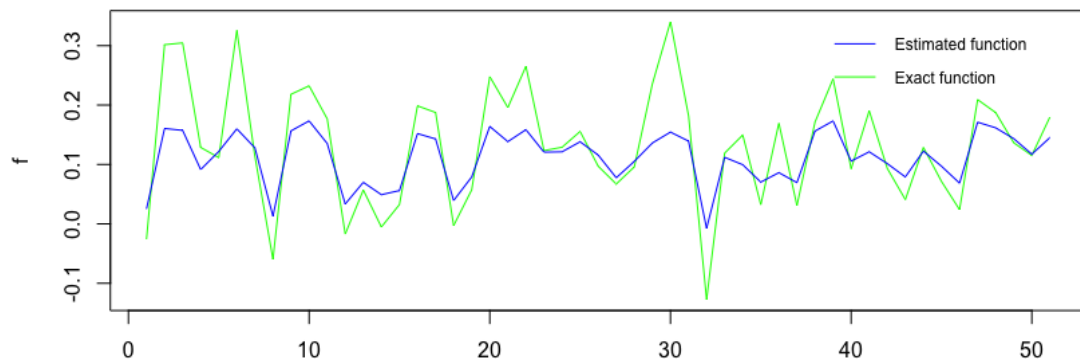


Figure 2.5: Regression function under exact and estimated function. The green line presents an exact function, which is many more variations, we desire to find the optimal parameters to obtain a more smooth curve. From Figure 2.4, we can chose these parameters following RMSE, however, in some cases it happens over-smoothing and under-smoothing. See in this Figure, the blue line looks like have some desirable properties when we chose the big enough values of  $\lambda$  and  $l$ .

Hence from the Figure 2.5 and Table 3.1, we can see that by choosing the optimal parameters for  $\gamma$ ,  $l$  and  $\lambda$  we can obtain a very well estimation function  $\hat{f}$  without under-smoothing or over-smoothing issues. Our regression model stay well above the RMSE criterion.

Our interpretation for these results is that, because of the nature of the simulated data  $(\mu_i, f(\mu_i))$  working directly on distributions and with the Wasserstein distance, is more appropriate than using linear projections. Indeed, in particular,



two distributions with similar means and small variances are close to each other with respect to both Wasserstein distance and the value of the output function  $f$ . However, the probability density functions of the two distributions are very different from each other with respect to the  $L^2$  distance in the case that the ratio between the two variances is large. Hence linear projections based on probability density functions is inappropriate in the setting considered here.

### 2.5.2 Simulation in the regression model with two-dimensional distribution input

In this section, we aim to study the two of estimated regression functions constructed by the Wasserstein distance and the Sliced Wasserstein distance in the case of 2D distribution entries.

#### Wasserstein distance and Sliced Wasserstein distance of 2D distributions

For the Wasserstein distance between two 2-dimensional Gaussian distributions, we refer the construction of positive definite kernels by Hilbert space embedding of optimal transport maps in [Bachoc et al. \(2018\)](#). Firstly, we introduce briefly the notion of Wasserstein barycenter. Let  $\mathbb{P}$  be a distribution in  $\mathcal{W}_2(\mathbb{W}_2(\mathbb{R}^p))$ , that is the set of measures on  $\mathcal{W}_2(\mathbb{R}^p)$  with finite expected variances, and consider  $\mu_1, \dots, \mu_n$  i.i.d probabilities drawn according to the distribution  $\mathbb{P}$ . In this frame work, the Wasserstein distance between distributions on  $\mathcal{W}_2(\mathbb{R}^p)$  is defined, for any  $\nu \in \mathcal{W}_2(\mathbb{R}^p)$ , as

$$W_2^2(\mathbb{P}, \delta_\nu) = \int W_2^2(\nu, \mu) d\mathbb{P}(\mu).$$

If  $\tilde{\mu}$  is a random distribution obeying law  $\mathbb{P}$ , this corresponds to

$$W_2^2(\mathbb{P}, \delta_\nu) = \mathbb{E}_{\tilde{\mu} \sim \mathbb{P}} W_2^2(\tilde{\mu}, \nu).$$

Note that we use the same notations for the Wasserstein distances over distributions in  $\mathcal{W}_2(\mathbb{R}^p)$  and over distributions on distributions in  $\mathcal{W}_2(\mathbb{W}_2(\mathbb{R}^p))$  because the space  $\mathcal{W}(\mathbb{W}_2(\mathbb{R}^p))$  inherits the properties of the space  $\mathcal{W}_2(\mathbb{R}^p)$ .

We define the Wasserstein barycenter of  $\mathbb{P}$  as a probability measure  $\bar{\mu}$  in  $\mathcal{W}_2(\mathbb{R}^p)$  such that

$$\int W_2^2(\bar{\mu}, \mu) d\mathbb{P}(\mu) = \inf \left\{ \int W_2^2(\nu, \mu) d\mathbb{P}(\mu), \nu \in \mathcal{W}_2(\mathbb{R}^p) \right\}.$$

The following theorem from [Alvarez-Esteban, del Barrio, Cuesta-Albertos, and Matrán \(2015\)](#) guarantees the existence and uniqueness of this barycenter under some assumptions.

**Theorem 2.11.** *Let  $\mathbb{P} \in \mathcal{W}_2(\mathbb{W}_2(\mathbb{R}^p))$ . Assume that every distribution in the support of  $\mathbb{P}$  is absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}^p$ . Then there exists a unique distribution  $\bar{\mu} \in \mathcal{W}_2(\mathbb{R}^p)$  defined as*

$$\bar{\mu} = \arg \min_{\nu \in \mathcal{W}_2(\mathbb{R}^p)} \left\{ \int W_2^2(\nu, \mu) d\mathbb{P}(\mu) \right\}.$$

Considering a reference distribution  $\bar{\mu} \in \mathcal{W}_2(\mathbb{R}^p)$ , for  $\mu \in \mathcal{W}_2(\mathbb{R}^p)$ , let  $T_\mu : \mathbb{R}^p \rightarrow \mathbb{R}^p$  we refer the optimal transportation maps defined by

$$T_{\mu\#}\mu = \bar{\mu},$$

where  $f_\# \pi = \pi \circ f^{-1}$  is the push-forward measure of a function  $f$  from a measure  $\pi$ , and

$$\|id - T_\mu\|_{L^2(\mu)} = W_2(\mu, \bar{\mu}).$$

Note that the map  $T_\mu$  is uniquely defined when  $\mu$  is absolutely continuous w.r.t Lebesgue measure. Furthermore,  $T_\mu$  is invertible from the support of  $\mu$  to the support of  $\bar{\mu}$  if also  $\bar{\mu}$  is absolutely continuous.

By associating the transport map  $T_\mu^{-1}$  to each distribution  $\mu$  and using positive definite kernel on the Hilbert  $L^2(\bar{\mu})$ , containing these transport maps, the groups of authors in [Bachoc et al. \(2018\)](#) introduce the following construction of positive definite kernels by Hilbert space embedding of optimal transport maps.

**Proposition 2.12.** *Consider a function  $F : \mathbb{R}^+ \rightarrow \mathbb{R}$  such that, for any Hilbert space  $H$  with norm  $\|\cdot\|_H$ , the function  $h_1, h_2 \rightarrow F(\|h_1 - h_2\|_H)$  is positive definite on  $H$ . Let  $\bar{\mu}$  be a continuous distribution in  $\mathcal{W}_2(\mathbb{R}^p)$ . Consider the function  $K$  on the set continuous distributions in  $\mathcal{W}_2(\mathbb{R}^p)$  defined by*

$$K(\mu, \nu) = F\left(\|T_\mu^{-1} - T_\nu^{-1}\|_{L^2(\bar{\mu})}\right). \quad (2.5.5)$$

Then  $K$  is positive definite.

Note that in the particular case of Gaussian distributions, the transport map is linear. Thus the kernel (2.5.5) based on the optimal transport map of Gaussian distributions is guaranteed universality. Using this result, we make later the link to our theory in building an estimated function for the unknown function  $f$  in our model corresponding to two-dimensional Gaussian distribution entries. In the general dimension of Gaussian distributions, we refer to the section 4.3 in [Bachoc et al. \(2018\)](#) for more details.

We introduce now the kernel obtained by the Sliced Wasserstein distance, the one built in [Kolouri et al. \(2015, 2018\)](#). The principal idea of this distance is to first obtain a family of one-dimensional probability distributions by using the Radon transforms. Then we calculate the distance between two multidimensional distributions as the Wasserstein distance of their one-dimensional representations. Let  $\nu$  and  $\mu$  be two continuous probability measures on  $\mathbb{R}^d$  with corresponding the positive probability density function  $I_\nu$  and  $I_\mu$ , the Sliced Wasserstein distance is defined by

$$K(\mu, \nu) = \exp\left(-\xi SW^2(\mu, \nu)\right),$$

where  $0 < \xi \leq 1$  and

$$SW^2(\mu, \nu) = \int_{\mathbb{S}^{d-1}} W_2^2(\mathcal{R}I_\mu(\cdot, \theta), \mathcal{R}I_\nu(\cdot, \theta)) d\theta,$$

with  $\mathcal{R}$  is the Radon transform and  $\mathbb{S}^{d-1}$  is the unit sphere in  $\mathbb{R}^d$ .

### Detail on the simulation procedure

We describe clearly our simulation in this section. We address the input-output map given by

$$f(\nu) = \frac{\|m_\nu\|_2}{0.05 + \|\Sigma_\nu\|_F},$$

where  $\nu$  is a two-dimensional Gaussian distribution of mean  $m_\nu$  and non-degenerated covariance  $\Sigma_\nu$ ,  $\|\cdot\|_F$  is the Frobenious norm.

We consider the predicted function  $\hat{f}$  as follows

$$\hat{f}(\nu) = \sum_{j=1}^n \hat{\alpha}_j K(\nu, \mu_j),$$

and we aim to compare the qualify of our estimated function  $\hat{f}$  corresponding to the Wasserstein distance and the one corresponds to the Sliced Wasserstein distance.

Particularly, the Wasserstein distance between two two-dimensional Gaussian distributions is defined by

$$W_2^2(\mathcal{N}_2(m_1, \Sigma_1), \mathcal{N}_2(m_2, \Sigma_2)) = \|m_1 - m_2\|_2^2 + \|\Sigma_1^{1/2} - \Sigma_2^{1/2}\|_F^2.$$

From that, for  $\mu = \mathcal{N}_2(m_\mu, \Sigma_\mu)$  and  $\nu = \mathcal{N}_2(m_\nu, \Sigma_\nu)$  we consider the kernel  $K$  such as

$$K(\mu, \nu) = \gamma^2 \exp\left(-\frac{\|m_\mu - m_\nu\|_2^2 + \|\Sigma_\mu^{1/2} - \Sigma_\nu^{1/2}\|_F^2}{l}\right),$$

where  $\gamma \neq 0$  and  $l > 0$ . We call this model by "Wasserstein2D".

For the Sliced Wasserstein distance, let  $\mu = \mathcal{N}_2(m_\mu, \Sigma_\mu)$ , a slice/projection of the Radon transform of  $\mu$  is then a one-dimensional Gaussian distribution  $\mathcal{R}\mu = \mathcal{N}(\theta m_\mu, \theta^T \Sigma_\mu \theta)$ , with  $\theta \in \mathbb{S}^1$ . Then we use the Wasserstein distance for two of one-dimensional Gaussian distributions. We are interested in the kernel defined in [Kolouri et al. \(2015\)](#), precisely

$$K(\mu, \nu) = \exp\left(-\xi \left(\|\theta m_\nu - \theta m_\mu\|_2^2 + \left\|(\theta^T \Sigma_\nu \theta)^{1/2} - (\theta^T \Sigma_\mu \theta)^{1/2}\right\|_F^2\right)\right), \quad (2.5.6)$$

where  $\xi > 0$ . We denote this model by "SlicedWasserstein2D".

Use the same technique in generating a testing and training set the one-dimensional Gaussian distribution cases. We consider the optimal estimates of both cases of "Wasserstein2D" and "SlicedWasserstein2D" regression models. More precisely, we optimize parameters  $\lambda > 0$ ,  $\gamma \neq 0$  and  $l > 0$  in "Wasserstein2D" model and  $\theta \in \mathbb{S}^1$  such as  $\theta = \left(\cos\left(\frac{c\pi}{50}\right), \sin\left(\frac{c\pi}{50}\right)\right)$  with  $c \in \{0, 1, \dots, 50\}$ ,  $\xi > 0$ ,  $\lambda > 0$  for "SlicedWasserstein2D" model. We consider the RMSE criteria for these models following the level of noise  $\sigma = 1, 5, 10$  in the Table 2.2.

Hence from the Table 2.2, we can see that by choosing the optimal parameters for  $\gamma, l, \lambda$  and  $\theta$  for obtaining the RMSE of both regression models, in the low noise level, the RMSE of both models are the same as quite good, but the "Wasserstein2D" model looks better than "SlicedWassertein2D" model when the noise is increased.

$\sigma$	"Wasserstein2D"	"SlicedWasserstein2D"
1	0.06	0.06
5	0.09	0.33
10	0.1	0.91

Table 2.2: RMSE values of the quality criteria for the "Wasserstein2D" and "SlicedWasserstein2D" regression models following the level of noise  $\sigma = 1, 5, 10$  with two-dimensional distribution entries. Basing on the minimal RMSE values, the "Wasserstein2D" model clearly outperforms the "SlicedWasserstein2D" model when there exists much noise in the regression model.

### 2.5.3 Application on evolution of hearing sensitivity

An otoacoustic emission (OAE) is a sound which is generated from within the inner ear. OAEs can be measured with a sensitive microphone in the ear canal and provide a noninvasive measure of cochlear amplification (see Chapter: Hearing basics in [Eggermont \(2017\)](#)). Recording of OAEs has become the main method for newborn and infant hearing screening (see Chapter: Early Diagnosis and Prevention of Hearing Loss in [Eggermont \(2017\)](#)). There are two types of OAEs: spontaneous otoacoustic emissions (SOAEs), which can occur without external stimulation, and evoked otoacoustic emissions (EOAEs), which require an evoking stimulus. In this paper, we consider a type of EOAEs that is Transient-EOAE (TEOAE) (see for instance in [Joris, Bergevin, Kalluri, Mc Laughlin, Michelet, van der Heijden, and Shera \(2011\)](#)), in which the evoked response from a click covers the frequency range up to around 4kHz. More precisely, each TEOAE models the ability of the cochlea to response to some frequencies in order to transform a sound into an information that will be processed by the brain. So to each observation is associated a curve (the Oto-Emission curve) which describes the response of the cochlea at several frequencies to a sound. The level of response depends on each individual and each stimulus should be normalized, but the way each individual reacts is characteristic of its physiological characteristic. Hence to each individual is associated a curve, which after normalization, it is considered as a distribution  $\mu$  describing the repartition of the responses for different frequencies ranging from 0 to 10 kHz. These distributions are shown in [Figure 2.6](#) and [Table 2.3](#).

The relationship between age and hearing sensitivity is investigated in [Collet, Moulin, Gartner, and Morgon \(1990\)](#); [O-Uchi, Kanzaki, Satoh, Yoshihara, Ogata, Inoue, and Mashino \(1994\)](#). The results show that when age increases, the presence of EOAEs by age group and the frequency peak in spectral analysis decreases and EOAE threshold increases. The differences in EOAE have been also reported between age classes in humans. These results convey the idea that the response evolves with age and that the effect of ages in hearing issues is deeply related to the changes of the cochlear properties. Hence our model uses as input these distributions  $\mu$  and

Name	Age	0(Hz)	39.06	78. 12	...	1171.88	1210.94	...	9765.62	9804.69
ABBAS	23	0	0.0006	0.0013	...	0.0819	0.0388	...	0.0021	0.0015
ADAMS	27	0.0001	0.0010	0.0022	...	0.0283	0.0283	...	0.0011	0.0006
ADENIYI	30	0.0002	0.0003	0.0014	...	0.0231	0.0065	...	0.0012	0.0016
DUPLOOY	17	0.0003	0.0005	0.0015	...	0.0786	0.1272	...	0.0036	0.0031
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
TRIMM	20	0.0005	0.0006	0.0026	...	0.0133	0.0215	...	0.0002	0.0017
VELE	26	0.0001	0.0005	0.0018	...	0.1176	0.0859	...	0.0003	0.0005
WALLER	40	0.0001	0.0001	0.0003	...	0.0178	0.0210	...	0.0014	0.0013
WILLIAM	22	0.0002	0.0003	0.0009	...	0.0156	0.0656	...	0.0014	0.0018

Table 2.3: TEOAE data. 48 individuals are considered in human population, recorded in South Africa, with last names in first column, their exact ages in second column and the others describe the responses of the cochlea at several frequencies ranging from 0Hz to 10kHz.

try to build a regression model to link between the age and these distributions representing the response of the cochlea at frequencies ranging from 0Hz to 10kHz. More precisely, we estimate the age for each level of response normalized and treated as a distribution  $\mu$  by using our proposed function as follows

$$\hat{f}(\mu_i) = \gamma^2 \sum_{j=1}^n \hat{\alpha}_j \exp \left( -\frac{\int_0^1 (F_{\mu_i}^{-1}(t) - F_{\mu_j}^{-1}(t))^2 dt}{l} \right), \quad (2.5.7)$$

where  $F_{\mu}^{-1}(U)$  defined as (2.3.4) and the value of  $\hat{\alpha}_j$  is chosen by optimal parameter  $\lambda$  in (2.4.9). We estimate the integral in (2.5.7) by following formula

$$\int_0^1 (F_{\mu_i}^{-1}(t) - F_{\mu_j}^{-1}(t))^2 dt = \sum_{m=1}^M \left[ F_{\mu_i}^{-1} \left( \frac{m}{M} \right) - F_{\mu_j}^{-1} \left( \frac{m}{M} \right) \right]^2, \quad (2.5.8)$$

where we can understand each  $F_{\mu_i}$  is an experimental distribution function of  $\mu_i$  and  $M$  is the number of discretized frequencies. As far as we know, each individual is associated with a curve, which after normalization without lost relationship among original data, it is considered as a distribution  $\mu_i$ . To calculate  $F_{\mu_i}^{-1} \left( \frac{m}{M} \right)$ , we arrange each curve in ascending order, for instance we denote  $X_{\mu}(1) \leq X_{\mu}(2) \leq \dots \leq X_{\mu}(M)$  following to distribution  $\mu$  and  $\sum_{m=1}^M X_{\mu}(m) = 1$ , so  $F_{\mu_i}^{-1} \left( \frac{m}{M} \right) = X_{\mu_i}(m)$ . Hence, we write again the formula (2.5.8)

$$\int_0^1 (F_{\mu_i}^{-1}(t) - F_{\mu_j}^{-1}(t))^2 dt = \sum_{m=1}^M (X_{\mu_i}(m) - X_{\mu_j}(m))^2, \quad (2.5.9)$$

where  $X_{\mu_i}$  is a curve  $X$  following to distribution  $\mu_i$ .

In our simulation, we choose  $\gamma = 1$ , the value of  $l = 10$  and the value of  $\lambda > 0$ . We aim to study the age in relation with its TEOAE curve of 48 subjects, recorded on human population in South Africa, with the range of frequency from 0Hz to 10kHz. See the Figure 2.7 to show the differences between the age of 15 to 50 years old. Following the estimated function in (2.5.7), we take 47 distributions  $\{\mu_j\}_{j=1}^{47}$

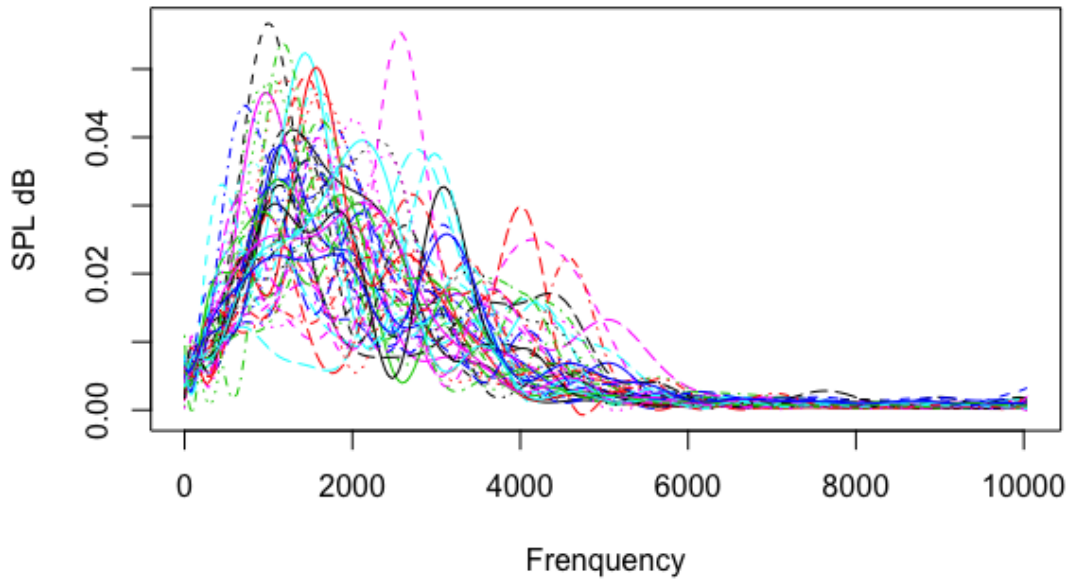


Figure 2.6: Oto-emission curves. 48 TEOAE curves following to frequencies ranging from 0Hz to 10kHz.

for training set to calculate estimation value of  $\hat{\alpha}_j$  and try to estimate real age of a remaining individual  $\mu_i$  with  $i \neq j$ . And the results are showed clearly in the Figure 2.8 and Figure 2.9 about the exact age and predicted age.

Hence in figure 2.7 and Figure 2.9, we applied effectively our proposed estimation function in predicting age from its TEOAE data. By choosing the optimal parameters  $\gamma$ ,  $l$  and  $\lambda$  we could predict very well the exact ages belonging to the age class [20, 30] and negligible errors in other age classes. This is quite reasonable when seeing in the Figure 2.7 that the age distributed diversity almost from 20 to 30 years old, so our proposed estimation function learnt very well to predict age in this age class. Thus by using the distribution regression model, we investigated the relationship between the evoked responses from clicks covering the frequencies range up to 10kHz and its evolutionary ages.

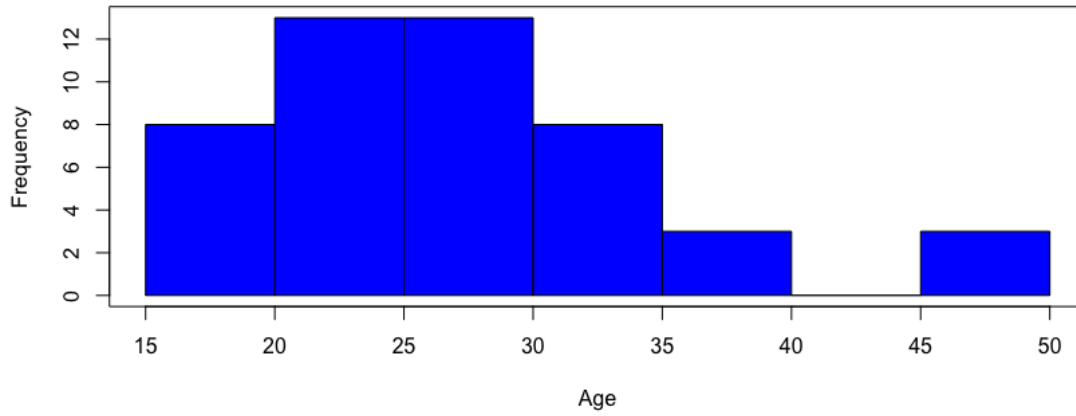


Figure 2.7: Histogram of real age in a human population. The age distribute diversity from 15 to 30, however, there is a few of individual of age from 35 to 40 and 45 to 50. And there exists no individual have age from 40 to 45.

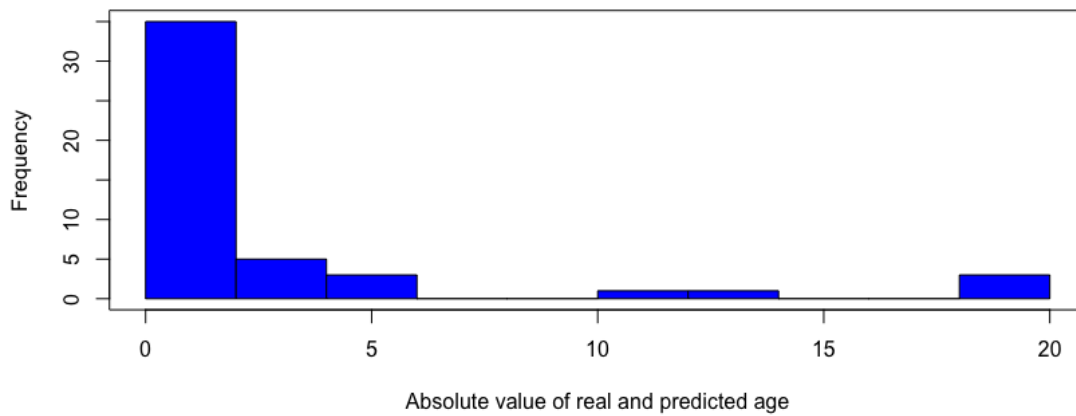


Figure 2.8: Histogram of difference between real and predicted age for OAE. In the first column, in which the difference between real and estimate age is very small closing to zero, this means more accuracy between real and predicted age. Almost the ages from 20 to 35 lie in this column.

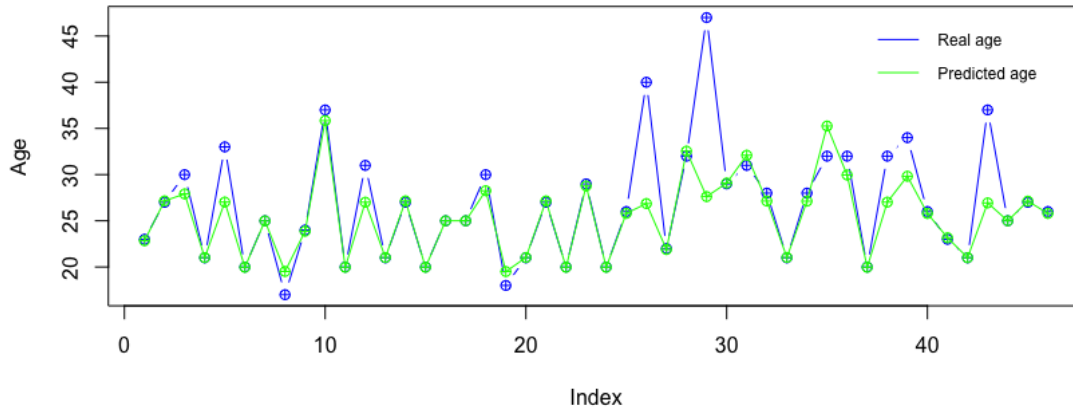


Figure 2.9: Real and Predicted Age. By using the optimized parameters of  $\gamma = 1$ ,  $l = 10$  and  $\lambda > 0$  depending on age class, we obtain almost the exacted ages belonging to the age class  $[20, 30]$  with  $\lambda$  around 15. For instance, we can predict very well the exact ages 20, 21, 23, 24, 25, 27, 29 corresponding to the predicted ages 19.50, 21.02, 22.83, 23.87, 24.89, 27.15, 28.76.

## 2.6 Discussion

In this paper, we have introduced a new estimated function for regression model with distribution inputs. More precisely, we effectively used class of positive definite kernel produced by Wasserstein distance, built in [Bachoc et al. \(2017\)](#) by proving that it is a kind of universal kernel. Researching the universal kernel theories, we detected a very good property of our universal kernel to build a RKHS. Then we obtained a particular estimation from Representer theorem for our distribution regression problem, these works showed that the relation between the random distribution and the real number response can be learnt by using directly the regularized empirical risk over RKHS. Our proposed estimation is clearly better than state-of-the-art-ones in simulated data. More interestingly, we researched successfully TEOAE curve of each individual in human population as a distribution which after normalization. We then investigated the relationship between age and its TEOAE that the response involves with age and the effect of age in hearing issues is deeply related to the change of cochlear. This is a new interesting approach in the field of Biostatistics, in which we indicated the evolution of hearing capacity under statistical domain - distribution regression model. We believe that our paper tackles an important issue for data science experts willing to predict problems in regression with probability distributions as input. The extension of this work on distributions for general dimensions should be addressed in a further work, using for instance as a kernel the one built in [Bachoc et al. \(2018\)](#).





## Part II

### Part 2: Testing hypothesis



# Chapitre 3

## Aggregated kernel based tests for signal detection in a regression model

The results of this chapter are represented in [Bui \(2019\)](#).

### Contents

---

<b>3.1</b>	<b>Introduction</b>	<b>64</b>
<b>3.2</b>	<b>Single tests based on a single kernel.</b>	<b>66</b>
3.2.1	Definition of the testing procedure.	66
3.2.2	Probabilities of first and second kind errors of the test.	68
3.2.3	Performance of the Monte Carlo approximation.	70
<b>3.3</b>	<b>Two particular examples of kernel function.</b>	<b>71</b>
3.3.1	Projection kernels.	71
3.3.2	Gaussian kernels.	73
<b>3.4</b>	<b>Multiple or aggregated tests based on collections of kernel functions.</b>	<b>74</b>
3.4.1	The aggregated testing procedure.	75
3.4.2	The aggregation of projection kernels.	75
3.4.3	The aggregation of Gaussian kernels.	77
<b>3.5</b>	<b>Simulation study.</b>	<b>78</b>
3.5.1	Presentation of the simulation study.	78
3.5.2	Simulation results.	79
<b>3.6</b>	<b>Appendices</b>	<b>81</b>
3.6.1	Proof of Proposition 3.1	81
3.6.2	Proof of Theorem 3.2	86
3.6.3	Proof of Corollary 3.5 and 3.7.	87

3.6.4	Proof of Proposition 3.6 . . . . .	87
3.6.5	Proof of Proposition 3.8 . . . . .	88
3.6.6	Proof of Theorem 3.9, Corollary 3.10 and 3.12. . . . .	89
3.6.7	Proof of Corollary 3.11. . . . .	90
3.6.8	Proof of Corollary 3.12. . . . .	90

Considering a regression model, we address the question of testing the nullity of the regression function. The testing procedure is available when the variance of the observations is unknown and does not depend on any prior information on the alternative. We first propose a single testing procedure based on a general symmetric kernel and an estimation of the variance of the observations. The corresponding critical values are constructed to obtain non asymptotic level- $\alpha$  tests. We then introduce an aggregation procedure to avoid the difficult choice of the kernel and of the parameters of the kernel. The multiple tests satisfy non-asymptotic properties and are adaptive in the minimax sense over several classes of regular alternatives.

### 3.1 Introduction

We observe  $(X_i, Y_i)_{1 \leq i \leq n}$  that obey to the regression model described as follows,

$$Y_i = f(X_i) + \sigma \epsilon_i, \quad i = 1, \dots, n. \quad (3.1.1)$$

We assume that  $X = (X_1, X_2, \dots, X_n)$  are i.i.d real random variables with values in a measurable set  $E$  such that  $[0, 1] \subset E \subset \mathbb{R}$  with bounded density  $\nu$  with respect to the Lebesgue measure on  $E$  and  $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$  are i.i.d standard Gaussian variables, independent of  $(X_1, X_2, \dots, X_n)$ . All along the paper,  $f$  is assumed to be in  $\mathbb{L}^2(E, d\nu)$ . We also assume that  $\|f\|_\infty = \sup_{x \in E} |f(x)| < +\infty$ . In order to estimate  $\sigma^2$ , we assume that we also observe  $(Y'_1, \dots, Y'_n)$  that obey to the model

$$Y'_i = f\left(\frac{i}{n}\right) + \sigma \epsilon'_i, \quad i = 1, \dots, n, \quad (3.1.2)$$

where  $\epsilon' = (\epsilon'_1, \dots, \epsilon'_n)$  is independent of  $(X_1, \dots, X_n, \epsilon_1, \dots, \epsilon_n)$ .

Given the observation of  $(X_i, Y_i)_{1 \leq i \leq n}$ ,  $(Y'_i)_{1 \leq i \leq n}$ , we want to test the null hypothesis

$$(H_0) : f = 0,$$

against the alternative

$$(H_1) : f \neq 0.$$

Hypothesis testing in nonparametric regression have been considered in the papers by [King \(1988\)](#), [Hardle and Marron \(1990\)](#), [Hall and Hart \(1990\)](#), [King et al. \(1991\)](#) and [Delgado \(1992\)](#). Tests for no effect in nonparametric regression are investigated in [Eubank and LaRiccia \(1993\)](#). In the paper of [Spokoiny et al. \(1996\)](#), the authors

considered the particular case where  $\sigma$  is assumed to be known. They propose tests that achieve the minimax rates of testing [up to an unavoidable  $\log \log(n)$  factor] for a wide range of Besov classes. [Baraud et al. \(2003\)](#) propose a test, based on model selection methods, for testing in a fixed design regression model that  $(f(X_1), \dots, f(X_n))$  belongs to a linear subspace of  $\mathbb{R}^n$  against a nonparametric alternative. They obtain optimal rates of testing up to a possible  $\log n$  factor over various classes of alternatives simultaneously. More recently, in a Poisson process framework, [Fromont et al. \(2012, 2013\)](#) consider two independent Poisson processes and address the question of testing equality of their respective intensities. They introduce tests based on a single kernel function and aggregate several kernel based tests to obtain adaptive minimax testing procedures over alternatives based on Besov or Sobolev balls.

In this work, we propose to construct aggregated kernel based testing procedures of  $(H_0)$  versus  $(H_1)$  in a regression model. Our test statistics are based on a single kernel function which can be chosen either as a projection or Gaussian kernel and we propose an estimation for the unknown variance  $\sigma^2$ . Our tests are exactly (and not only asymptotically) of level  $\alpha$ . We obtain the optimal non-asymptotic conditions on the alternative which guarantee that the probability of second kind error is at most equal to a prescribed level  $\beta$ . However, the testing procedures that we introduce hereafter also intend to overcome the question of calibrating the choice of kernel and/or the parameters of the kernel. They are based on an aggregation approach, that is well-known in adaptive testing ([Baraud et al. \(2003\)](#) and [Fromont et al. \(2013\)](#)). This paper is strongly inspired by the paper of [Fromont et al. \(2013\)](#). Instead of considering a particular single kernel, we consider a collection of kernels and the corresponding collection of tests, each with an adapted level of significance. We then reject the null hypothesis when there exists at least one of the tests in the collection which rejects the null hypothesis. The aggregated testing procedures are constructed to be of level  $\alpha$  and the loss in second kind error due to the aggregation, when unavoidable, is as small as possible. Then we prove that these multiple tests satisfy the adaptive minimax properties over several classes of alternatives. At last, we compare our tests with tests investigated in [Eubank and LaRiccia \(1993\)](#) from a practical point of view.

The paper is organized as follows. We describe the single tests based on a single kernel function with the corresponding critical values approximated by a Monte Carlo method in Section 3.2. In Section 3.3, we specify the performances of the single tests for two particular examples of kernels and explain the reasons why we need to aggregate tests based on a collection of kernel functions which are presented in Section 3.4. We present the simulation study in Section 3.5 and the major proofs are given in Appendix.

## 3.2 Single tests based on a single kernel.

### 3.2.1 Definition of the testing procedure.

We assume that we observe  $\{(X_i, Y_i)\}_{i=1}^n$  that obey to model (3.1.1). In order to estimate the unknown variance  $\sigma^2$ , we assume that we observe another sample  $(Y'_i)_{1 \leq i \leq n}$  from the model (3.1.2). We are interested in testing the null hypothesis  $(H_0) : f = 0$  against  $(H_1) : f \neq 0$ . Let  $K$  be a symmetric kernel function:  $E \times E \rightarrow \mathbb{R}$  satisfying:

**Assumption 1.**

$$\int_{E^2} K^2(x, y) f(x) f(y) d\nu(x) d\nu(y) < +\infty.$$

We introduce the test statistic  $V_K$  defined as follows,

$$V_K = \frac{T_K}{\hat{\sigma}_n^2}, \quad (3.2.1)$$

where

$$T_K = \frac{1}{n(n-1)} \sum_{i \neq j=1}^n K(X_i, X_j) Y_i Y_j \quad (3.2.2)$$

and

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n/2} (Y'_{2i-1} - Y'_{2i})^2, \quad (3.2.3)$$

where for the sake of simplicity, we assume that  $n$  is even. Let us now introduce some notations. We set  $K_{ij} = K(X_i, X_j)$ ,  $f_i = f(X_i)$  and  $C(a, b)$  is a constant depending on  $a$  and  $b$ , that will be used all along the paper and may vary from line to line.

The expectation of  $T_K$  is equal to

$$\begin{aligned} \mathbb{E}[T_K] &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{1}{n(n-1)} \sum_{i \neq j=1}^n K_{ij} (f(X_i) + \sigma \epsilon_i) (f(X_j) + \sigma \epsilon_j) \middle| X \right] \right] \\ &= \mathbb{E} \left[ \frac{1}{n(n-1)} \sum_{i \neq j=1}^n K_{ij} f(X_i) f(X_j) \right] \\ &= \int_{E^2} K(x, y) f(x) f(y) d\nu(x) d\nu(y). \end{aligned}$$

In the following, we denote for all  $x \in E$ ,

$$K[f](x) = \int_E K(x, y) f(y) d\nu(y),$$

and for all  $f, g \in \mathbb{L}^2(E, d\nu)$

$$\langle f, g \rangle = \int_E f(x) g(x) d\nu(x) \text{ and } \|f\|^2 = \langle f, f \rangle.$$

Within these notations,

$$\mathbb{E}(T_K) = \langle K[f], f \rangle, \quad (3.2.4)$$

whose existence is ensured by Assumption 1. We now compute the expectation of  $\widehat{\sigma}_n^2$ .

$$\begin{aligned} \mathbb{E}[\widehat{\sigma}_n^2] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n/2} \left[ \left( f\left(\frac{2i-1}{n}\right) + \sigma \epsilon'_{2i-1} \right) - \left( f\left(\frac{2i}{n}\right) + \sigma \epsilon'_{2i} \right) \right]^2\right] \\ &= \frac{1}{n} \sum_{i=1}^{n/2} \left[ f\left(\frac{2i-1}{n}\right) - f\left(\frac{2i}{n}\right) \right]^2 + \frac{1}{n} \sum_{i=1}^{n/2} \sigma^2 \mathbb{E}(\epsilon'_{2i-1} - \epsilon'_{2i})^2 \\ &= a^2 + \sigma^2, \end{aligned}$$

with  $a^2 := \frac{1}{n} \sum_{i=1}^{n/2} \left[ f\left(\frac{2i-1}{n}\right) - f\left(\frac{2i}{n}\right) \right]^2$ .

Thus  $\widehat{\sigma}_n^2$  is a biased estimator of  $\sigma^2$  with bias  $a^2$ . If  $f$  is a regular function this bias will be small.

We have chosen to consider and study in this paper two possible examples of kernel functions. For each example, we give a simpler expression of  $\mathbb{E}(T_K)$ .

*Example 1.* When  $E = [0, 1]$ , our first choice for  $K$  is a symmetric kernel function based on a finite orthonormal family  $\{\phi_\lambda, \lambda \in \Lambda\}$  with respect to the scalar product  $\langle \cdot, \cdot \rangle$ ,

$$K(x, y) = \sum_{\lambda \in \Lambda} \phi_\lambda(x) \phi_\lambda(y). \quad (3.2.5)$$

For all  $f$  in  $\mathbb{L}^2([0, 1], d\nu)$  we get

$$\begin{aligned} K[f](x) &= \int_0^1 \left( \sum_{\lambda \in \Lambda} \phi_\lambda(x) \phi_\lambda(y) \right) f(y) d\nu(y) \\ &= \sum_{\lambda \in \Lambda} \left( \int_0^1 \phi_\lambda(y) f(y) d\nu(y) \right) \phi_\lambda(x) = \Pi_S(f), \end{aligned}$$

where  $S$  is the subspace of  $\mathbb{L}^2([0, 1], d\nu)$  generated by the functions  $\{\phi_\lambda, \lambda \in \Lambda\}$  and  $\Pi_S$  denotes the orthogonal projection onto  $S$  for  $\langle \cdot, \cdot \rangle$ . Thus

$$\mathbb{E}(T_K) = \langle \Pi_S(f), f \rangle.$$

Hence, when  $\{\phi_\lambda, \lambda \in \Lambda\}$  is well-chosen,  $T_K$  can also be viewed as a relevant estimator of  $\|f\|^2$ .

*Example 2.* When  $E = \mathbb{R}$  and  $\nu$  is a density function respect to the Lebesgue measure on  $\mathbb{R}$ , our second choice for  $K$  is a Gaussian kernel defined by,

$$K(x, y) = \frac{1}{h} k\left(\frac{x-y}{h}\right), \text{ for all } (x, y) \in \mathbb{R}^2 \quad (3.2.6)$$



where  $k(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$ , for all  $u \in \mathbb{R}$  and  $h$  is a positive bandwidth. Then, for all  $f \in \mathbb{L}^2(\mathbb{R}, d\nu)$  we have

$$K[f](x) = \int_{-\infty}^{\infty} \frac{1}{h} k\left(\frac{x-y}{h}\right) f(y) d\nu(y) = k_h * f(x),$$

where  $*$  is the convolution producer with respect to the measure  $\nu$  and  $k_h(u) = \frac{1}{h} k\left(\frac{u}{h}\right)$ ,  $\forall u \in \mathbb{R}$ . Thus in this case

$$\mathbb{E}(T_K) = \langle k_h * f, f \rangle.$$

Hence, when the bandwidth  $h$  is well chosen,  $T_K$  can also be viewed as a relevant estimator of  $\|f\|^2$ .

From the choices of the two examples above for  $K$ , we have seen that the test statistic  $V_K$  can be viewed as a relevant estimator of  $\|f\|^2$ . Thus, it seems to be reasonable proposal to consider a test which rejects  $(H_0)$  when  $V_K$  is as "large enough". Now, we define the critical values used in our tests.

We define

$$V_K^{(0)} = \frac{\frac{1}{n(n-1)} \sum_{i \neq j=1}^n K(X_i, X_j) \epsilon_i \epsilon_j}{\frac{1}{n} \sum_{i=1}^{n/2} (\epsilon'_{2i-1} - \epsilon'_{2i})^2}. \quad (3.2.7)$$

Note that, under  $(H_0)$ , conditionally on  $X$ ,  $V_K$  and  $V_K^{(0)}$  have exactly the same distribution. We now choose the quantile of the conditional distribution of  $V_K^{(0)}$  given  $X$  as the critical value for our test. This quantity can easily be estimated by simulations.

More precisely, for  $\alpha$  in  $(0, 1)$ , if  $q_{K,1-\alpha}^{(X)}$  denotes the  $(1 - \alpha)$  quantile of the distribution of  $V_K^{(0)}$  conditionally on  $X$ , we consider the test that rejects  $(H_0)$  when  $V_K > q_{K,1-\alpha}^{(X)}$ . The corresponding test function is defined by

$$\Phi_{K,\alpha} = \mathbb{1}\{V_K > q_{K,1-\alpha}^{(X)}\}. \quad (3.2.8)$$

Notice that in practice, the true quantile  $q_{K,1-\alpha}^{(X)}$  is not available, but it can be approximated by a Monte Carlo procedure.

### 3.2.2 Probabilities of first and second kind errors of the test.

Since under  $(H_0)$ ,  $V_K$  and  $V_K^{(0)}$  have the same distribution conditionally on  $X$ , for any  $\alpha \in (0, 1)$ , we have

$$\mathbb{P}_{(H_0)} \left( V_K > q_{K,1-\alpha}^{(X)} \middle| X \right) \leq \alpha.$$

By taking the expectation over  $X$ , we obtain

$$\mathbb{P}_{(H_0)} (\Phi_{K,\alpha} = 1) \leq \alpha.$$

Let us now consider an alternative hypothesis, corresponding to a non zero regression function  $f$ . Given  $\beta$  in  $(0, 1)$ , we now aim to determine a non-asymptotic condition on the regression function  $f$  which guarantees that  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$ . Denoting by  $q_{K,1-\beta/2}^\alpha$  the  $(1 - \beta/2)$  quantile of the conditional quantile  $q_{K,1-\alpha}^{(X)}$ ,

$$\begin{aligned} \mathbb{P}_f(\Phi_{K,\alpha} = 0) &= \mathbb{P}_f(V_K \leq q_{K,1-\beta/2}^\alpha) + \mathbb{P}_f(V_K \leq q_{K,1-\alpha}^{(X)}, q_{K,1-\alpha}^{(X)} > q_{K,1-\beta/2}^\alpha) \\ &\leq \mathbb{P}_f(V_K \leq q_{K,1-\beta/2}^\alpha) + \beta/2. \end{aligned}$$

Thus, a condition which guarantees that  $\mathbb{P}_f(V_K \leq q_{K,1-\beta/2}^\alpha) \leq \beta/2$  will ensure that  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$ . The following proposition gives such a condition.

**Proposition 3.1.** *Let  $\alpha, \beta$  be the fixed levels in  $(0, 1)$ . We have that*

$$\mathbb{P}_f(V_K \leq q_{K,1-\beta/2}^\alpha) \leq \beta/2,$$

as soon as

$$\langle K[f], f \rangle \geq \sqrt{\frac{16A_K + 8B_K}{\beta}} + D_{n,\beta} q_{K,1-\beta/2}^\alpha, \quad (3.2.9)$$

with

$$\begin{aligned} A_K &= \frac{n-2}{n(n-1)} \int_E (K[f](x))^2 [f^2(x) + \sigma^2] d\nu(x), \\ B_K &= \frac{1}{n(n-1)} \int_{E^2} K^2(x, y) [f^2(x) + \sigma^2] [f^2(y) + \sigma^2] d\nu(x)d\nu(y), \\ D_{n,\beta} &= \sigma^2 + a^2 + \frac{4\sigma^2}{n} \sqrt{\left(\frac{n}{2} + \frac{na^2}{\sigma^2}\right) \ln\left(\frac{2}{\beta}\right) + \frac{4\sigma^2}{n} \ln\left(\frac{2}{\beta}\right)}. \end{aligned}$$

Thus we have, under (3.2.9),

$$\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta.$$

Moreover, there exists some constant  $\kappa > 0$  such that, for every  $K$  and  $n \geq 32 \ln(2/\alpha)$

$$q_{K,1-\beta/2}^\alpha \leq \frac{2\kappa}{\sqrt{n(n-1)}} \ln\left(\frac{2}{\alpha}\right) \sqrt{\frac{2 \int_{E^2} K^2(x, y) d\nu(x)d\nu(y)}{\beta}}. \quad (3.2.10)$$

To prove the first part of this result, we simply use Markov's inequality for the term  $T_K$  and an exponential inequality for non-central Chi-square variables due to (Birgé (2001)) for the term  $\hat{\sigma}_n^2$ . The control of  $q_{K,1-\beta/2}^\alpha$  derives from a property of Gaussian chaos combined with an exponential inequality (due to De la Pena and Giné (2012) and Huskova and Janssen (1993)). The detailed proof is given in the Appendix.

The following theorem gives a condition on  $\|f\|^2$  for the test to be powerful.

**Theorem 3.2.** *Let  $\alpha, \beta$  be fixed levels in  $(0, 1)$ ,  $\kappa$  be a positive constant,  $K$  be a symmetric kernel function, and  $\Phi_{K,\alpha}$  be the test defined by (3.2.8). Let  $C_K$  be an upper bound for  $\int_{E^2} K^2(x, y) d\nu(x) d\nu(y)$ . Then for all  $n \geq 32 \ln(2/\alpha)$ , we have  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$ , as soon as*

$$\begin{aligned} \|f\|^2 &\geq \|f - K[f]\|^2 + \frac{16 (\|f\|_\infty^2 + \sigma^2)}{n\beta} \\ &+ \frac{4}{\sqrt{n(n-1)\beta}} \left( \kappa D_{n,\beta} \ln\left(\frac{2}{\alpha}\right) + \sqrt{2} (\|f\|_\infty^2 + \sigma^2) \right) \sqrt{C_K}. \end{aligned} \quad (3.2.11)$$

The right hand side of the above inequality corresponds to a bias-variance trade-off. For particular choices of the kernel function  $K$ , these terms will be upper bounded in Section 3.3.

### 3.2.3 Performance of the Monte Carlo approximation.

In this section, we introduce a Monte Carlo method used to approximate the conditional quantiles  $q_{K,1-\alpha}^{(X)}$  by  $\hat{q}_{K,1-\alpha}^{(X)}$  as follows. We consider the set of  $2B$  independent sequences of i.i.d standard Gaussian variables

$$\{\epsilon^b, 1 \leq b \leq B\} \quad \text{and} \quad \{\epsilon'^b, 1 \leq b \leq B\},$$

where  $\epsilon^b = \{\epsilon_i^b\}_{i=1}^n$ ,  $\epsilon'^b = \{\epsilon'_i{}^b\}_{i=1}^n$ ,  $1 \leq b \leq B$ .

We define

$$V_K^{(\epsilon^b, \epsilon'^b)} = \frac{\frac{1}{n(n-1)} \sum_{i \neq j=1}^n K(X_i, X_j) \epsilon_i^b \epsilon_j^b}{\frac{1}{n} \sum_{i=1}^{n/2} (\epsilon'_{2i-1}{}^b - \epsilon'_{2i}{}^b)^2},$$

where  $X = (X_1, \dots, X_n)$  are observed from model (3.1.2).

Under  $(H_0)$ , conditionally on  $X$ , the variables  $V_K^{(\epsilon^b, \epsilon'^b)}$  have the same distribution function as  $V_K$  and as  $V_K^{(0)}$ . We denote by  $F_{K,B}$  the empirical distribution function of the sample  $\left\{ V_K^{(\epsilon^b, \epsilon'^b)}, 1 \leq b \leq B \right\}$ , conditionally on  $X$ .

$$\forall x \in \mathbb{R}, \quad F_{K,B}(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1} \left\{ V_K^{(\epsilon^b, \epsilon'^b)} \leq x \right\}.$$

Then the Monte Carlo approximation of  $q_{K,1-\alpha}^{(X)}$  is defined by

$$\hat{q}_{K,1-\alpha}^{(X)} = F_{K,B}^{-1}(\alpha) = \inf \{ t \in \mathbb{R}, F_{K,B}(t) \geq 1 - \alpha \}.$$

We recall the test function defined in (3.2.8) and we reject  $(H_0)$  when  $V_K > q_{K,1-\alpha}^{(X)}$  with  $q_{K,1-\alpha}^{(X)}$  the  $(1 - \alpha)$  quantile of  $V_K^{(0)}$  defined by (3.2.7) conditionally on  $X$ . Now, by using the estimated quantile  $\hat{q}_{K,1-\alpha}^{(X)}$ , we consider the test given by

$$\hat{\Phi}_{K,\alpha} = \mathbb{1} \left\{ V_K > \hat{q}_{K,1-\alpha}^{(X)} \right\}. \quad (3.2.12)$$

For the test defined in (3.2.12), the probabilities of first and second kind errors can be upper bounded. This is the purpose of the two following propositions, whose proofs are given in Fromont et al. (2013).

**Proposition 3.3.** *Let  $\alpha$  be some fixed level in  $(0, 1)$ , and  $\widehat{\Phi}_{K,\alpha}$  be the test defined by (3.2.12). Then,*

$$\mathbb{P}_{(H_0)} \left( \widehat{\Phi}_{K,\alpha} = 1 \mid X \right) \leq \frac{B\alpha + 1}{B + 1}.$$

**Proposition 3.4.** *Let  $\alpha$  and  $\beta$  be fixed levels in  $(0, 1)$  such that  $\alpha_B = \alpha - \sqrt{\ln B / (2B)}$  and  $\beta_B = \beta - 2/B > 0$ . Let  $\widehat{\Phi}_{K,\alpha}$  be the test given in (3.2.12). Let  $A_K, B_K, D_{n,\beta}$  and  $\kappa$  as in Proposition 3.1, and let  $q_{K,1-\beta_B/2}^{\alpha_B}$  be the  $(1-\beta_B/2)$  quantile of  $q_{K,1-\alpha_B}^{(X)}$ . If*

$$\langle K[f], f \rangle > \sqrt{\frac{16A_K + 8B_K}{\beta}} + D_{n,\beta_B} q_{K,1-\beta_B/2}^{\alpha_B}, \quad (3.2.13)$$

then  $\mathbb{P}_f \left( \widehat{\Phi}_{K,\alpha} = 0 \right) \leq \beta$ . Moreover,

$$q_{K,1-\beta_B/2}^{\alpha_B} \leq \frac{2\kappa}{\sqrt{n(n-1)}} \ln \left( \frac{2}{\alpha_B} \right) \sqrt{\frac{2 \int_{E^2} K^2(x, y) d\nu(x) d\nu(y)}{\beta_B}}. \quad (3.2.14)$$

*Comments.* When comparing (3.2.13) and (3.2.14) with (3.2.9) and (3.2.10) in Proposition 3.1, we notice that they asymptotically coincide when  $B \rightarrow +\infty$ . Moreover, if  $\alpha = \beta = 0.05$  and  $B \geq 6000$ , the multiplicative factor of  $\frac{\kappa}{\sqrt{n(n-1)}} \sqrt{\int_{E^2} K^2(x, y) d\nu(x) d\nu(y)}$  is of order 1.2 in (3.2.14) compared with (3.2.10).

### 3.3 Two particular examples of kernel function.

In this section, we specify the performances of the above test for two examples of the kernels including projection kernels and Gaussian kernels.

#### 3.3.1 Projection kernels.

We assume  $E = [0, 1]$ . We consider the projection kernel defined in (3.2.5) and aim to give a more explicit formulation for the result of Theorem 3.2 under the choice of this kernel. We also evaluate the uniform separation rates over Besov bodies.

**Corollary 3.5.** *Let  $\alpha, \beta \in (0, 1)$  and  $\kappa > 0$  be a constant. Let  $\Phi_{K,\alpha}$  be defined in (3.2.8), where  $K$  is the projection kernel defined by (3.2.5). We denote by  $S$  the linear subspace of  $\mathbb{L}^2([0, 1], d\nu)$ , generated by the functions  $\{\phi_\lambda, \lambda \in \Lambda\}$ , and we*

assume that the dimension of  $S$  is equal  $D$ . Then  $n \geq 32 \ln(\alpha/2)$  if

$$\begin{aligned} \|f\|^2 &\geq \|f - \Pi_S(f)\|^2 + \frac{16(\|f\|_\infty^2 + \sigma^2)}{n\beta} \\ &+ \frac{4\sqrt{D}}{\sqrt{n(n-1)\beta}} \left( \kappa D_{n,\beta} \ln\left(\frac{2}{\alpha}\right) + \sqrt{2}(\|f\|_\infty^2 + \sigma^2) \right), \end{aligned}$$

then

$$\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta.$$

Let us consider the particular case when the kernel  $K$  is the projection kernel onto the space generated by functions of the Haar basis defined as follows.

Let  $\{\phi_0, \phi_{(j,k)}, j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}\}$  be the Haar basis of  $\mathbb{L}^2([0, 1])$  with

$$\phi_0(x) = \mathbb{1}_{[0,1]}(x) \quad \text{and} \quad \phi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \quad (3.3.1)$$

where  $\psi(x) = \mathbb{1}_{[0,1/2)}(x) - \mathbb{1}_{[1/2,1]}(x)$ . The linear subspace  $S$  is generated by a subset of the Haar basis. More precisely, we denote by  $S_0$  the subspace of  $\mathbb{L}^2([0, 1])$  generated by  $\phi_0$ , and we define

$$K_0(x, x') = \phi_0(x)\phi_0(x'). \quad (3.3.2)$$

We also consider, for  $J \geq 1$  the subspace  $S_J$  generated by  $\{\phi_\lambda, \lambda \in \{0\} \cup \Lambda_J\}$  with  $\Lambda_J = \{(j, k), j \in \{0, \dots, J-1\}, k \in \{0, \dots, 2^j - 1\}\}$ , and

$$K_J(x, x') = \sum_{\lambda \in \{0\} \cup \Lambda_J} \phi_\lambda(x)\phi_\lambda(x'). \quad (3.3.3)$$

We set  $\alpha_0 = \langle f, \phi_0 \rangle$  and for every  $j \in \mathbb{N}$ ,  $k \in \{0, \dots, 2^j - 1\}$ ,  $\alpha_{j,k} = \langle f, \phi_{j,k} \rangle$ .

We now introduce the Besov body defined for  $\delta > 0$ ,  $R > 0$  by

$$\mathcal{B}_{2,\infty}^\delta(R) = \left\{ f \in \mathbb{L}^2([0, 1], d\nu), f = \alpha_0 \phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha_{j,k} \phi_{j,k} / \alpha_0^2 \leq R^2, \forall j \in \mathbb{N}, \sum_{k=0}^{2^j-1} \alpha_{j,k}^2 \leq R^2 2^{-2j\delta} \right\}.$$

For all  $J \geq 0$ , we consider the kernel function  $K_J$  defined by (3.3.2), (3.3.3) and the associated test function  $\Phi_{K_J,\alpha}$  defined in (3.2.8) with  $K = K_J$ . For an optimal choice of  $J$ , realizing a good compromise between the bias term and the variance term appearing in (3.2.11), we give a condition of  $\|f\|^2$  for  $f \in \mathcal{B}_{2,\infty}^\delta(R)$  which ensures that the power of our test is larger than  $1 - \beta$ .

**Proposition 3.6.** *Let  $\alpha, \beta \in (0, 1)$ . For all  $J \geq 0$ , let  $K_J$  defined by (3.3.2), (3.3.3) and consider the test function  $\Phi_{K_{J^*},\alpha} = \mathbb{1}\{V_{K_{J^*}} > q_{K_{J^*},1-\alpha}^{(X)}\}$  where*

$$J^* = \left\lceil \log_2 \left( n^{2/(1+4\delta)} \right) \right\rceil. \quad (3.3.4)$$

For all  $f \in \mathcal{B}_{2,\infty}^\delta(\mathbb{R})$  such that

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, \|f\|_\infty) n^{-4\delta/(1+4\delta)}, \quad (3.3.5)$$

we have  $\mathbb{P}_f(\Phi_{K_{J^*}, \alpha} = 0) \leq \beta$ .

*Comments.*

1. Non asymptotic lower bounds for the rates of testing in signal detection over Besov bodies are given in [Baraud et al. \(2002\)](#). These lower bounds coincide with the bound given in (3.3.5), hence our result is sharp.
2. In (3.3.4),  $J^*$  depends on  $\delta$ , the regularity parameter of the Besov body, so it leads to the natural question of the choice of this parameter. In order to propose a procedure that is adaptive with respect to the regularity of the unknown regression function  $f$ , we introduce aggregated tests in Section 3.4.

### 3.3.2 Gaussian kernels.

For this second example, we assume that  $E = \mathbb{R}$ . We consider the Gaussian kernel defined in (3.2.6) and rewrite the result of Theorem 3.2 under the choice of this kernel. We also evaluate the uniform separation rates over Sobolev balls for this test.

**Corollary 3.7.** *Let  $\alpha, \beta \in (0, 1)$ ,  $\kappa > 0$  be a constant and  $\Phi_{K,\alpha}$  be the test function defined in (3.2.8) where  $K$  is defined in (3.2.6). For  $n \geq 32 \ln(2/\alpha)$  if*

$$\begin{aligned} \|f\|^2 \geq & \|f - k_h * f\|^2 + \frac{16 (\|f\|_\infty^2 + \sigma^2)}{n\beta} \\ & + \frac{4 \|\nu\|_\infty}{(2\pi)^{1/4} \sqrt{n(n-1)} \beta h} \left( \kappa D_{n,\beta} \ln \left( \frac{2}{\alpha} \right) + \sqrt{2} (\|f\|_\infty^2 + \sigma^2) \right). \end{aligned} \quad (3.3.6)$$

We obtain that

$$\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta.$$

Let  $E = \mathbb{R}$  and  $\mathcal{L} = \mathbb{N}^*$ . For  $x, y$  in  $\mathbb{R}$  and  $h = 2^{-l}$ , for all  $l \in \mathcal{L}$ , we consider

$$K_l(x, y) = \frac{1}{2^{-l}} k \left( \frac{x-y}{2^{-l}} \right), \quad (3.3.7)$$

with

$$k(u) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{u^2}{2} \right).$$

Let us introduce for  $\delta > 0$  the Sobolev ball  $\mathcal{S}_\delta(R)$  defined by

$$\mathcal{S}_\delta(R) = \left\{ s : \mathbb{R} \rightarrow \mathbb{R} \mid s \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |u|^{2\delta} |\hat{s}(u)|^2 du \leq 2\pi R^2 \right\},$$

where  $\hat{s}$  denotes the Fourier transform of  $s$ :  $\hat{s}(u) = \int_{\mathbb{R}} s(x)e^{i\langle x,u \rangle} dx$ .

For all  $l \in \mathcal{L}$ , we consider the kernel function  $K_l$  defined by (3.3.7) and the associated test function  $\Phi_{K_l, \alpha}$  defined in (3.2.8) with  $K = K_l$ . For an optimal choice of  $l$ , realizing a good compromise between the bias term and the variance term appearing in (3.3.6), we give a condition of  $\|f\|^2$  for  $f \in \mathcal{S}_\delta(R)$  which ensures that the power of our test is larger than  $1 - \beta$ .

**Proposition 3.8.** *Let  $\alpha, \beta \in (0, 1)$ . For all  $l \in \mathcal{L}$ , let  $K_l$  defined by (3.3.7) and the test function  $\Phi_{K_l, \alpha} = \mathbb{1}\{V_{K_l} > q_{K_l, 1-\alpha}^{(X)}\}$  we set*

$$l^* = \left\lceil \log_2 \left( n^{\frac{2}{1+4\delta}} \right) \right\rceil. \quad (3.3.8)$$

For all  $f \in \mathcal{S}_\delta(R)$  such that

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, \|f\|_\infty) n^{-4\delta/(1+4\delta)}. \quad (3.3.9)$$

We have  $\mathbb{P}_f(\Phi_{K_{l^*}, \alpha} = 0) \leq \beta$ .

*Comments.*

1. As in Proposition 3.6, we obtain in the right hand term of (3.3.9) a classical bound for the separation rates of testing over regular classes of alternatives such as Holderian balls (see Ingster (1993)) for nonparametric minimax rates of testing in various setups.
2. Non asymptotic lower bounds for the rates of testing in signal detection over Sobolev balls are given in Fromont and Lévy-Leduc (2006). These bounds coincide with the bound given in (3.3.9).
3. In (3.3.8), as previously,  $l^*$  depends on  $\delta$ , the regularity parameter of the Sobolev ball, so it leads to the natural question of the choice of this parameter answered through the aggregated tests in Section 3.4.

### 3.4 Multiple or aggregated tests based on collections of kernel functions.

In the previous section, we have considered testing procedures based on a single kernel function  $K$ . However, the following question is natural: how can we choose the kernel, and its parameters. For instance, the orthonormal family in the projection kernel in Section 3.3.1, the bandwidth  $h$  in the Gaussian kernel in Section 3.3.2. Baraud et al. (2003) proposed adaptive testing procedures based on the aggregation of a collection of tests. This idea is presented in a series of papers, among which Fromont et al. (2013) proposed an aggregation procedure. Following this idea, we consider in this section a collection of kernel functions instead of a single one. Beside that, we define a multiple testing procedure by aggregating the corresponding single tests, with an adapted choice of the critical values.

### 3.4.1 The aggregated testing procedure.

Let us describe the aggregated testing procedure by introducing a finite collection  $\{K_m, m \in \mathcal{M}\}$  of symmetric kernel functions:  $E \times E \rightarrow \mathbb{R}$ . For  $m \in \mathcal{M}$ , we replace  $K$  in (3.2.1) and (3.2.7) by  $K_m$  to define  $V_{K_m}$  and  $V_{K_m}^{(0)}$  and let  $\{w_m, m \in \mathcal{M}\}$  be a collection of positive numbers such that  $\sum_{m \in \mathcal{M}} e^{-w_m} \leq 1$ . Conditionally on  $X$ , for  $u \in (0, 1)$ , we denote by  $q_{m,1-u}^{(X)}$  the  $(1-u)$  quantile of  $V_{K_m}^{(0)}$ . Given  $\alpha$  in  $(0, 1)$ , we consider the test which rejects  $(H_0)$  when there exists at least one  $m$  in  $\mathcal{M}$  such that

$$V_{K_m} > q_{m,1-u_\alpha}^{(X)} e^{-w_m},$$

where  $u_\alpha^{(X)}$  is defined by

$$u_\alpha^{(X)} = \sup \left\{ u > 0, \mathbb{P} \left( \sup_{m \in \mathcal{M}} \left( V_{K_m} - q_{m,1-u}^{(X)} e^{-w_m} \right) > 0 \middle| X \right) \leq \alpha \right\}. \quad (3.4.1)$$

We consider the test function  $\Phi_\alpha$  defined by

$$\Phi_\alpha = \mathbb{1} \left\{ \sup_{m \in \mathcal{M}} \left( V_{K_m} - q_{m,1-u_\alpha}^{(X)} e^{-w_m} \right) > 0 \right\}. \quad (3.4.2)$$

Using the Monte Carlo method, we can estimate  $u_\alpha^{(X)}$  and the quantiles  $q_{m,1-u_\alpha}^{(X)} e^{-w_m}$  for all  $m \in \mathcal{M}$ . The following theorem provides a control of the first and second kind error for the test  $\Phi_\alpha$ . The detailed proof is given in the Appendix.

**Theorem 3.9.** *Let  $\alpha, \beta$  be fixed levels in  $(0, 1)$  and  $\Phi_\alpha$  be the test defined by (3.4.2). We have*

$$\mathbb{P}_{(H_0)} (\Phi_\alpha = 1) \leq \alpha. \quad (3.4.3)$$

And for all regression function  $f$ , we have

$$\mathbb{P}_f (\Phi_\alpha = 0) \leq \beta, \quad (3.4.4)$$

as soon as there exists  $m$  in  $\mathcal{M}$  such that

$$\mathbb{P} \left( V_{K_m} \leq q_{K_m,1-\alpha e^{-w_m}}^{(X)} \right) \leq \beta.$$

*Comments.* This theorem shows that the aggregated test is of level  $\alpha$ , for all  $m$ . Moreover, as soon as the second kind error is controlled by  $\beta$  for at least one test in the collection, the same holds for the aggregated procedure with the price that the level  $\alpha$  is replaced by  $\alpha e^{-w_m}$  to guarantee that the aggregated procedure is of level  $\alpha$ .

### 3.4.2 The aggregation of projection kernels.

Let us specify the performance of the aggregated test for a collection of projection kernels.



**Corollary 3.10.** *Let  $\alpha, \beta$  be fixed levels in  $(0, 1)$ . Let  $\{S_m, m \in \mathcal{M}\}$  be a finite collection of linear subspaces of  $\mathbb{L}^2([0, 1], d\nu)$ , generated by the functions  $\{\phi_\lambda, \lambda \in \Lambda_m\}$  and we assume that the dimension of  $S_m$  is equal to  $D_m$ . We set, for all  $m \in \mathcal{M}$ ,  $K_m(x, y) = \sum_{\lambda \in \Lambda_m} \phi_\lambda(x)\phi_\lambda(y)$ . Let  $\Phi_\alpha$  be defined by (3.4.2) with the collection of kernels  $\{K_m, m \in \mathcal{M}\}$  and the collection  $\{w_m, m \in \mathcal{M}\}$  of positive numbers such that  $\sum_{m \in \mathcal{M}} e^{-w_m} \leq 1$ .*

*Then  $\Phi_\alpha$  is a level  $\alpha$  test. Moreover,  $\mathbb{P}_f(\Phi_\alpha = 0) \leq \beta$  if*

$$\|f\|^2 \geq \inf_{m \in \mathcal{M}} \left\{ \|f - \Pi_{S_m}(f)\|^2 + \frac{16(\|f\|_\infty^2 + \sigma^2)}{n\beta} + \frac{4\sqrt{D_m}}{\sqrt{n(n-1)\beta}} \left( \kappa D_{n,\beta} \left( \ln\left(\frac{2}{\alpha}\right) + w_m \right) + \sqrt{2}(\|f\|_\infty^2 + \sigma^2) \right) \right\}, \quad (3.4.5)$$

where  $\kappa > 0$  and  $n \geq 32 \ln(\alpha/2)$ .

*Comments.* Comparing this result with the one obtained in Corollary 3.5 for the single test based on a projection kernel, we can see that the multiple testing procedure allows to obtain the infimum over all  $m$  in  $\mathcal{M}$  in the right hand side of (3.4.5) at the price of the additional term  $w_m$ .

Let us consider the particular case when the collection of kernels  $\{K_m, m \in \mathcal{M}\}$  is the collection of projection kernels based on the constructions in Section 3.3.1. Let for some  $\bar{J} \geq 1$ ,  $\mathcal{M}_{\bar{J}} = \{J, 0 \leq J \leq \bar{J}\}$ , and for all  $J$  in  $\mathcal{M}_{\bar{J}}$ ,  $w_J = 2(\ln(J+1) + \ln(\pi/\sqrt{6}))$ .

We consider  $\Phi_\alpha^{(1)}$ , the test defined by (3.4.2) with the collection of kernels  $\{K_J, 0 \leq J \leq \bar{J}\}$  where  $K_0, K_J, 0 < J \leq \bar{J}$  defined in (3.3.2), (3.3.3). We obtain from the Corollary 3.10 that there exists some constant  $C(\alpha, \beta, \sigma, \|f\|_\infty)$  such that  $\mathbb{P}_f(\Phi_\alpha^{(1)} = 0) \leq \beta$  as soon as

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, \|f\|_\infty) \inf_{J \in \mathcal{M}_{\bar{J}}} \left\{ \|f - \Pi_{S_J}(f)\|^2 + \ln(J+2) \frac{2^{J/2}}{n} \right\}. \quad (3.4.6)$$

For any  $\delta > 0, R, R' > 0$  we consider

$$\mathcal{B}_{2,\infty}^\delta(R, R') = \left\{ f : f \in \mathcal{B}_{2,\infty}^\delta(R), \|f\|_\infty \leq R' \right\}. \quad (3.4.7)$$

**Corollary 3.11.** *Let  $\alpha, \beta \in (0, 1)$ . For all  $J \in \mathcal{M}_{\bar{J}}$ , we consider the test function  $\Phi_\alpha^{(1)}$ . Assuming that  $\ln \ln(n) \geq 1$ ,  $2^{\bar{J}} \geq n^2$ . Then, for any  $\delta, R, R' > 0$  we set*

$$J^{**} = \left\lceil \log_2 \left( \left( \frac{n}{\ln \ln(n)} \right)^{\frac{2}{4\delta+1}} \right) \right\rceil.$$

*For all  $f \in \mathcal{B}_{2,\infty}^\delta(R, R')$  such that*

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, R') \left( \frac{\ln \ln(n)}{n} \right)^{\frac{4\delta}{4\delta+1}}, \quad (3.4.8)$$

we have  $\mathbb{P}_f(\Phi_\alpha^{(1)} = 0) \leq \beta$ .

*Comments.* We obtain a right hand term in (3.4.8) of order  $(\ln \ln(n)/n)^{4\delta/(1+4\delta)}$ . This rate of testing was shown to be optimal for the signal detection in a Gaussian white noise by Spokoiny et al. (1996). In particular, he showed that the logarithm factor is the price to pay for adaptation.

### 3.4.3 The aggregation of Gaussian kernels.

We here consider the aggregated test based on a collection of Gaussian kernels.

**Corollary 3.12.** *Let  $\alpha, \beta \in (0, 1)$ ,  $\{h_l, l \in \mathcal{L}\}$  be a collection of positive bandwidths, we consider  $\{K_l, l \in \mathcal{L}\}$  a collection of Gaussian kernels corresponding to the above collection of positive bandwidths, where  $K_l$  defined in (3.3.7). Let  $\Phi_\alpha$  be defined by (3.4.2) with the collection of kernel  $\{K_l, l \in \mathcal{L}\}$  and a collection  $\{w_l, l \in \mathcal{L}\}$  of positive numbers such that  $\sum_{l \in \mathcal{L}} e^{-w_l} \leq 1$ .*

*Then  $\Phi_\alpha$  is a level  $\alpha$  test. Moreover, there exists  $\kappa > 0$  such that if*

$$\|f\|^2 \geq \inf_{l \in \mathcal{L}} \left\{ \|f - k_l * f\|^2 + \frac{16(\|f\|_\infty^2 + \sigma^2)}{n\beta} + \frac{4\|\nu\|_\infty}{(2\pi)^{1/4}\sqrt{n(n-1)\beta h_l}} \left( \kappa D_{n,\beta} \left( \ln \left( \frac{2}{\alpha} \right) + w_l \right) + \sqrt{2}(\|f\|_\infty^2 + \sigma^2) \right) \right\}, \quad (3.4.9)$$

*We obtain that  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$ .*

For  $l \in \mathcal{L} = \mathbb{N} \setminus \{0\}$ . We consider the particular case where we take  $h_l = 2^{-l}$  and  $w_l = 2(\ln(l+1) + \ln(\pi^2/6))$  for all  $l \in \mathcal{L}$ . Let  $\Phi_\alpha^{(2)}$  be the test defined by (3.4.2) with the collection of Gaussian kernels  $\{K_l, l \in \mathcal{L}\}$  and  $\{w_l, l \in \mathcal{L}\}$ . We obtain from Corollary 3.12 that there exists  $C(\alpha, \beta, \sigma, \|f\|_\infty)$  such that  $\mathbb{P}_f(\Phi_\alpha^{(2)} = 0) \leq \beta$  if

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, \|f\|_\infty) \inf_{l \in \mathcal{L}} \left\{ \|f - k_l * f\|^2 + \frac{w_l}{n\sqrt{2^{-l}}} \right\}. \quad (3.4.10)$$

For  $\delta > 0, R, R' > 0$  we consider

$$\mathcal{S}_\delta(R, R') = \left\{ f : f \in \mathcal{S}_\delta(R), \|f\|_\infty \leq R' \right\}. \quad (3.4.11)$$

**Corollary 3.13.** *Let  $\alpha, \beta \in (0, 1)$ . For all  $l \in \mathcal{L}$ , we consider the test function  $\Phi_\alpha^{(2)}$  and assume that  $\ln \ln \geq 1$ . For any  $\delta > 0, R, R' > 0$ , we set*

$$l^{**} = \left\lceil \log_2 \left( \left( \frac{n}{\ln \ln(n)} \right)^{\frac{2}{4\delta+1}} \right) \right\rceil.$$

For all  $f \in \mathcal{S}_\delta(R, R')$  such that

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, R') \left( \frac{\ln \ln(n)}{n} \right)^{\frac{4\delta}{4\delta+1}},$$

we have  $\mathbb{P}_f(\Phi_\alpha^{(2)} = 0) \leq \beta$ .

*Comments.* The rate of testing is of order  $(\ln \ln(n)/n)^{4\delta/(1+4\delta)}$ . This rate was shown to be optimal over periodic Sobolev balls up to the logarithm, by [Castillo et al. \(2006\)](#).

## 3.5 Simulation study.

### 3.5.1 Presentation of the simulation study.

We study our aggregated testing procedures from a practical point of view in this section. We consider  $E = [0, 1]$ ,  $n = 100$  and choose  $\alpha = 0.05$ . In the following simulation,  $X_1, \dots, X_n$  are i.i.d uniform random variables on  $[0, 1]$ .

Let us introduce the collection of symmetric kernel functions and the aggregated testing procedure  $\Phi_\alpha$  defined by (3.4.2) as follows. First, we consider the test  $\Phi_\alpha^{(1)}$  denoted by P corresponding to a collection of projection kernels. To be more explicit, we consider the Haar basis  $\{\phi_0, \phi_{(j,k)}, j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}\}$  introduced in Section 3.3.1. Let  $K_0(x, x') = \phi_0(x)\phi_0(x')$  and for  $J \geq 1$   $K_J(x, x') = \sum_{\lambda \in \{0\} \cup \Lambda_J} \phi_\lambda(x)\phi_\lambda(x')$  with  $\Lambda_J = \{(j, k), j \in \{0, \dots, J-1\}, k \in \{0, \dots, 2^j - 1\}\}$ . Let  $\mathcal{M}_{\bar{J}} = \{J, 0 \leq J \leq 7\}$  and for all  $J$  in  $\mathcal{M}_{\bar{J}}$ ,  $w_J = 2(\ln(J+1) + \ln(\pi/\sqrt{6}))$ . We consider  $\Phi_\alpha^{(1)}$  the multiple testing procedure with the collection of kernels  $\{K_J, J \in \mathcal{M}_{\bar{J}}\}$ .

Second, we also consider the multiple test associated with the collection of Gaussian kernel functions defined in Section 3.4.3. For  $\mathcal{L} = \{1, 2, \dots, 6\}$  we take  $\{h_l, l \in \mathcal{L}\} = \{1/24, 1/16, 1/12, 1/8, 1/4, 1/2\}$ , let  $K_l(x, y) = \frac{1}{h_l} k\left(\frac{x-y}{h_l}\right)$  with  $k(u) = (2\pi)^{-1/2} \exp(-u^2/2)$ . Then taking  $w_l = 1/|\mathcal{L}| = 1/6$ , we consider  $\Phi_\alpha^{(2)}$  the multiple testing procedure denoted by G, with the collection of kernels  $\{K_l, l \in \mathcal{L}\}$ .

At last, we are interested in the collection of both projection and Gaussian kernels. We define  $\Phi_\alpha^{(3)}$  denoted by PG, the multiple testing procedure with the collection of kernels  $\{K_p, p \in \mathcal{P} = \mathcal{M}_{\bar{J}} \cup \mathcal{L}\}$ . For  $p \in \mathcal{M}_{\bar{J}}$  we take  $w_p = \ln(J+1) + \ln(\pi/\sqrt{6})$  and for  $p \in \mathcal{L}$  we take  $w_p = 1/12$ .

We recall that the test rejects  $(H_0)$  when there exists at least one  $m$  in  $\mathcal{M}$  such that  $V_{K_m} > q_{m, 1-u_\alpha^{(X)}}^{(X)} e^{-w_m}$ . Hence, for each observation  $X = (X_1, \dots, X_n)$  we have to estimate  $u_\alpha^{(X)}$  defined by (3.4.1) and  $q_{m, 1-u_\alpha^{(X)}}^{(X)}$ . Applying the Monte Carlo method introduced in the Section 3.2.3, these quantities are well approximated. To be more explicit, we generate 400000 samples of  $\{\epsilon^b\}_{b=1}^{400000}$  and  $\{\epsilon'^b\}_{b'=1}^{400000}$ , in

	1st error	CI
P	0.0504	[0.033, 0.068]
G	0.0506	[0.032, 0.068]
PG	0.0498	[0.032, 0.0657]

Table 3.1: The probabilities of first kind error of the test for  $\alpha = 0.05$  and the upper and lower bounds of an asymptotic confidence interval with confidence level 99%.

which we use one half to approximate the conditional probability occurring in (3.4.1) and other half is used to estimate the distribution of each  $V_{K_m}^{(0)}$ . We note that  $u_\alpha^{(X)}$  is approximated by the bisection method such that the estimated conditional probabilities in (3.4.1) are as close as possible to  $\alpha$ .

### 3.5.2 Simulation results.

We first study the probabilities of first kind error of each test. We realize 5000 simulations of  $X$ . For each simulation, we determine the conclusions of tests P, G and PG where the critical values are approximated by the Monte Carlo methods described above. The probabilities of first kind error of tests are estimated by the number of rejections for these tests divided by 5000. The obtained estimated levels of tests and the corresponding confidence intervals (CI) are showed in the Table 3.1.

We then study the probabilities of rejection for each test under several alternatives. We first consider the following alternative,

$$f_{1,a,\epsilon}(x) = \epsilon \mathbb{1}_{[0,a)}(x) - \epsilon \mathbb{1}_{[a,2a)}(x),$$

with  $0 < \epsilon \leq 1$  and  $0 < a < 1$ . Second, we consider the alternative defined by

$$f_{2,\tau}(x) = \tau \sum_j \frac{h_j}{2} (1 + \operatorname{sgn}(x - p_j)),$$

with  $\tau > 0$ , and  $h_j \in \mathcal{Z}$ ,  $0 < p_j < 1$  for all  $j$ . Next, we consider the following alternative,

$$f_{3,c}(x) = c \cos(10\pi x),$$

with  $c > 0$ . The last alternative, for which we aim to compare our results with the results of [Eubank and LaRiccia \(1993\)](#) is defined as follows

$$f_{4,\varrho,j}(x) = \varrho \cos(2\pi jx),$$

where  $\varrho \geq 0$  and  $j \in \mathbb{N}^*$ .

For each alternative  $f$ , we realize 1000 simulations of  $X$ . For each simulation, we determine conclusions of tests P, G and PG, where the critical values of our tests are still approximated by the Monte Carlo method. The powers of tests are estimated by the number of rejections divided by 1000. The obtained estimated powers of tests and lower bounds of the asymptotic confidence intervals with the confidence

$(a, \epsilon)$	(1/4, 0.7)		(1/4, 0.9)		(1/4, 1)		(1/8, 1)	
	$\hat{p}$	CI	$\hat{p}$	CI	$\hat{p}$	CI	$\hat{p}$	CI
P	0.876	[0.849, 0.903]	0.986	[0.976, 0.996]	0.996	[0.990, 1.001]	0.699	[0.662, 0.736]
G	0.831	[0.801, 0.861]	0.977	[0.965, 0.989]	0.992	[0.985, 0.999]	0.635	[0.596, 0.674]
PG	0.884	[0.858, 0.910]	0.984	[0.973, 0.994]	0.996	[0.991, 1.001]	0.690	[0.652, 0.727]

Table 3.2: The power of the test for the alternative  $f_{1,a,\epsilon}$  corresponding to  $(a, \epsilon) = (1/4, 0.7), (1/4, 0.9), (1/4, 1), (1/8, 1)$  and the upper and lower bounds of an asymptotic confidence intervals with confidence level 99%.

$\tau$	0.05		0.1		0.5	
	$\hat{p}$	CI	$\hat{p}$	CI	$\hat{p}$	CI
P	0.218	[0.177, 0.243]	0.654	[0.615, 0.693]	1	*
G	0.208	[0.175, 0.241]	0.668	[0.629, 0.704]	1	*
PG	0.210	[0.177, 0.243]	0.678	[0.639, 0.716]	1	*

Table 3.3: The power of the test for the alternative  $f_{2,\tau}$  corresponding to  $\tau = 1, 2, 3$  and the upper and lower bounds of an asymptotic confidence intervals with confidence level 99%.

$c$	1		2		3	
	$\hat{p}$	ICI	$\hat{p}$	CI	$\hat{p}$	CI
P	0.35	[0.311, 0.389]	0.90	[0.876, 0.924]	0.98	[0.969, 0.991]
G	0.56	[0.519, 0.600]	0.98	[0.967, 0.991]	1	*
PG	0.34	[0.301, 0.379]	0.89	[0.864, 0.915]	1	*

Table 3.4: The power of the test for the alternative  $f_{3,c}$  corresponding to  $c = 1, 2, 3$  and the upper and lower bounds of an asymptotic confidence intervals with confidence level 99%.

level 99% are represented in the Table 3.2, 3.3 and 3.4. Table 3.5 is proposed for comparing our tests and the two of tests  $T_{nm}$  denoted by EL1,  $T_{n\lambda}$  denoted by EL2, which were proposed in Eubank and LaRiccia (1993). We recall briefly tests  $T_{nm}$ ,  $T_{n\lambda}$  as follows.

$$T_{nm} = \frac{\sum'_{|j| \leq m} |\tilde{a}_{jn}|^2 - 2m\sigma^2}{2\sigma^2\sqrt{m}},$$

and

$$T_{n\lambda} = \frac{n \sum'_{|j| \leq (n-1)/2} |a_{jn}|^2 (1 + \lambda(2\pi j)^4)^{-2} - 2\sigma^2 \sum'_{|j| \leq (n-1)/2} (1 + \lambda(2\pi j)^4)^{-2}}{2\sigma^2 \left( \sum'_{|j| \leq (n-1)/2} (1 + \lambda(2\pi j)^4)^{-4} \right)^{1/2}},$$

where  $\sum'$  indicates summation excluding the zero index and  $\tilde{a}_{jn}$  are the sample Fourier coefficients,

$$\tilde{a}_{jn} = \frac{1}{n} \sum_{i=1}^n y_i e^{-2\pi j(i-1)/n}.$$

In the three alternatives  $f_{1,a,\epsilon}$ ,  $f_{2,\tau}$  and  $f_{3,c}$ , the test PG is more powerful than P and G tests. Our conclusion is that the test PG is a good choice in practice. In

	Test	$\varrho = 0$	$\varrho = 0.5$	$\varrho = 1$	$\varrho = 1.5$
$j = 1$	P	0.049	0.606	1	1
	G	0.048	0.459	0.99	1
	PG	0.048	0.441	0.99	1
	EL1	0.074	0.837	1	1
	EL2	0.062	0.805	1	1
$j = 3$	P	0.053	0.224	0.905	1
	G	0.053	0.630	0.922	1
	PG	0.049	0.228	1	1
	EL1	0.069	0.718	1	1
	EL2	0.058	0.693	1	1
$j = 6$	P	0.043	0.134	0.696	0.990
	G	0.044	0.146	0.741	0.995
	PG	0.045	0.134	0.700	0.996
	EL1	0.076	0.134	0.428	0.979
	EL2	0.056	0.107	0.368	0.961

Table 3.5: The power of the test for the alternative  $f_{4,\varrho,j}$  corresponding to  $\varrho = 0, 0.5, 1, 1.5$ ,  $j = 1, 2, 3$ .

Table 3.5, we see in the first column ( $\rho = 0$ ), which corresponds to the null hypothesis that our test is of level  $\alpha = 0.05$ , which is not the case for the tests proposed by [Eubank and LaRiccia \(1993\)](#), which are only asymptotically of level  $\alpha$ . This explains why our test is generally less powerful than the tests EL1 and EL2 for  $\rho = 0.5$ . In the other cases, we obtain as good or better results.

## 3.6 Appendices

### 3.6.1 Proof of Proposition 3.1

Let us prove the first part of Prop 3.1. Recall that  $q_{K,1-\beta/2}^\alpha$  denotes the  $(1 - \beta/2)$  quantile of  $q_{K,1-\alpha}^{(X)}$  which is the  $(1 - \alpha)$  quantile of  $V_K^{(0)}$  conditionally on  $X$ . We here want to find a condition on  $\varepsilon_K = \mathbb{E}(T_K)$ , ensuring that

$$\mathbb{P}_f(V_K \leq q_{K,1-\beta/2}^\alpha) \leq \beta/2.$$

From Markov's inequality, we have for all  $\lambda > 0$

$$\mathbb{P}_f(|-T_K + \varepsilon_K| \geq \lambda) \leq \frac{\mathbb{E}(T_K^2) - \varepsilon_K^2}{\lambda^2}. \quad (3.6.1)$$

Let us compute  $\mathbb{E}(T_K^2|X)$ . We see that

$$T_K^2 = \frac{1}{n^2(n-1)^2} \left[ \sum_{\substack{i \neq j \neq u \neq v \\ i,j,u,v=1}}^n K_{ij}K_{uv}Y_iY_jY_uY_v + 4 \sum_{\substack{i \neq j \neq u \\ i,j,u=1}}^n K_{ij}K_{iu}Y_i^2Y_jY_u + 2 \sum_{\substack{i \neq j \\ i,j=1}}^n K_{ij}^2Y_i^2Y_j^2 \right].$$

Then

$$\begin{aligned}\mathbb{E} [T_K^2 | X] &= \frac{1}{n^2(n-1)^2} \sum_{i \neq j \neq u \neq v=1}^n K(X_i, X_j) K(X_u, X_v) f(X_i) f(X_j) f(X_u) f(X_v) \\ &+ \frac{4}{n^2(n-1)^2} \sum_{i \neq j \neq u=1}^n K(X_i, X_j) K(X_i, X_u) [f^2(X_i) + \sigma^2] f(X_j) f(X_u) \\ &+ \frac{2}{n^2(n-1)^2} \sum_{i \neq j=1}^n K^2(X_i, X_j) [f^2(X_i) + \sigma^2] [f^2(X_j) + \sigma^2]\end{aligned}$$

Since  $\mathbb{E} [T_K^2] = \mathbb{E} [\mathbb{E} [T_K^2 | X]]$ , and since  $(X_1, \dots, X_n)$  are i.i.d with density  $\nu$  on  $E$ , we obtain

$$\begin{aligned}\mathbb{E} [T_K^2] &= \frac{(n-2)(n-3)}{n(n-1)} \int_{E^4} K(x, y) K(u, v) f(x) f(y) f(u) f(v) d\nu(x) d\nu(y) d\nu(u) d\nu(v) \\ &+ \frac{4(n-2)}{n(n-1)} \int_{E^3} K(x, y) K(x, u) [f^2(x) + \sigma^2] f(y) f(u) d\nu(x) d\nu(y) d\nu(u) \\ &+ \frac{2}{n(n-1)} \int_{E^2} K^2(x, y) [f^2(x) + \sigma^2] [f^2(y) + \sigma^2] d\nu(x) d\nu(y) \\ &= \frac{(n-2)(n-3)}{n(n-1)} \varepsilon_K^2 + 4A_K + 2B_K.\end{aligned}$$

Thus

$$\mathbb{E} [T_K^2] - \frac{(n-2)(n-3)}{n(n-1)} \varepsilon_K^2 = 4A_K + 2B_K. \quad (3.6.2)$$

In fact

$$0 < \frac{(n-2)(n-3)}{n(n-1)} < 1, \quad \forall n > 3.$$

Then

$$\mathbb{E} [T_K^2] - \frac{(n-2)(n-3)}{n(n-1)} \varepsilon_K^2 \geq \mathbb{E} [T_K^2] - \varepsilon_K^2.$$

Replacing (3.6.2) into (3.6.1) we obtain

$$\mathbb{P}_f (|\varepsilon_K - T_K| \geq \lambda) \leq \frac{4A_K + 2B_K}{\lambda^2}. \quad (3.6.3)$$

Choosing  $\lambda = \sqrt{\frac{16A_K + 8B_K}{\beta}}$ , the above inequality leads to

$$\mathbb{P}_f \left( |\varepsilon_K - T_K| \geq \sqrt{\frac{16A_K + 8B_K}{\beta}} \right) \leq \frac{\beta}{4}.$$

This implies

$$\mathbb{P}_f \left( T_K \leq \varepsilon_K - \sqrt{\frac{16A_K + 8B_K}{\beta}} \right) \leq \frac{\beta}{4}. \quad (3.6.4)$$

Now we consider the term  $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (Y_{2i-1} - Y_{2i})^2$ . Following to the Cochran's theorem, we consider the orthogonal subspace  $W$  of dimension  $n/2$ . We denote

$(e_1, \dots, e_{n/2})$  be an orthogonal basis of  $W$ , where for all  $i = 1, \dots, n/2$ ,  $e_i^T$  is a vector includes  $n$  elements within two values  $\{0, 1\}$  and its values equal to 1 at two positions  $2i$  and  $2i - 1$ . On the other hand, for  $Y = (Y_1, \dots, Y_n)$ , with  $Y_i = f(X_i) + \sigma\epsilon_i$  we have

$$Y = \mathcal{N}_n \left( \begin{pmatrix} f(X_1) \\ \vdots \\ f(X_n) \end{pmatrix}, \sigma^2 I_n \right).$$

Using the Cochran's theorem, we have

$$\|\Pi_{W^\perp} Y\|^2 = \|Y - \Pi_W Y\|^2 \sim \sigma^2 \chi^2 \left( \frac{n}{2}, \frac{na^2}{2\sigma^2} \right),$$

where  $a^2 := \frac{1}{n} \sum_{i=1}^{n/2} \left[ f\left(\frac{2i-1}{n}\right) - f\left(\frac{2i}{n}\right) \right]^2$  and  $\chi^2(k, \lambda)$  denotes a non central Chi-square variable with  $k$  degrees of freedom and non centrality parameter  $\lambda$ .

Moreover,

$$\begin{aligned} \|Y - \Pi_W Y\|^2 &= \sum_{i=1}^{n/2} \left( Y_{2i-1} - \frac{Y_{2i-1} + Y_{2i}}{2} \right)^2 + \left( Y_{2i} - \frac{Y_{2i-1} + Y_{2i}}{2} \right)^2 \\ &= 2 \sum_{i=1}^{n/2} \left( \frac{Y_{2i-1} - Y_{2i}}{2} \right)^2 = \frac{1}{2} \sum_{i=1}^{n/2} (Y_{2i-1} - Y_{2i})^2 \\ &= \frac{n}{2} \hat{\sigma}_n^2. \end{aligned}$$

Hence

$$\hat{\sigma}_n^2 \sim \frac{2\sigma^2}{n} \chi^2 \left( \frac{n}{2}, \frac{na^2}{2\sigma^2} \right).$$

Now, we consider the variable  $Z \sim \chi^2 \left( \frac{n}{2}, \frac{na^2}{2\sigma^2} \right)$ . Using Lemma 8.1 in [Birgé \(2001\)](#), we have

$$\forall \rho > 0, \mathbb{P}_f \left[ Z \geq \frac{n}{2} + \frac{na^2}{2\sigma^2} + 2\sqrt{\left( \frac{n}{2} + \frac{na^2}{\sigma^2} \right) \rho} + 2\rho \right] \leq e^{-\rho}.$$

This implies

$$\forall \rho > 0, \mathbb{P}_f \left[ \hat{\sigma}_n^2 \geq \sigma^2 + a^2 + \frac{4\sigma^2}{n} \sqrt{\left( \frac{n}{2} + \frac{na^2}{\sigma^2} \right) \rho} + \frac{4\rho\sigma^2}{n} \right] \leq e^{-\rho}. \quad (3.6.5)$$

Choosing  $\rho = \ln(4/\beta)$ , (3.6.5) leads to

$$\mathbb{P}_f \left( \hat{\sigma}_n^2 \geq D_{n,\beta} \right) \leq \frac{\beta}{4},$$

where

$$D_{n,\beta} := \sigma^2 + a^2 + \frac{4\sigma^2}{n} \sqrt{\left( \frac{n}{2} + \frac{na^2}{\sigma^2} \right) \ln \left( \frac{4}{\beta} \right)} + \frac{4\sigma^2}{n} \ln \left( \frac{4}{\beta} \right).$$



Thus

$$\mathbb{P}_f \left( \frac{1}{\hat{\sigma}_n^2} \leq \frac{1}{D_{n,\beta}} \right) \leq \frac{\beta}{4}. \quad (3.6.6)$$

From (3.6.4) and (3.6.6), we obtain

$$\mathbb{P}_f (V_K \leq UV) \leq \beta/2, \quad (3.6.7)$$

with

$$U = \varepsilon_K - \sqrt{\frac{16A_K + 8B_K}{\beta}},$$

$$V = \frac{1}{D_{n,\beta}}.$$

If  $q_{K,1-\beta/2}^\alpha \leq UV$  we have  $\mathbb{P}_f (V_K \leq q_{K,1-\beta/2}^\alpha) \leq \beta/2$ .  
Therefore, if

$$\varepsilon_K \geq D_{n,\beta} q_{K,1-\beta/2}^\alpha + \sqrt{\frac{16A_K + 8B_K}{\beta}}, \quad (3.6.8)$$

then

$$\mathbb{P}_f (\Phi_{K,\alpha} = 0) \leq \beta.$$

Let us now give an upper bound for  $q_{K,1-\beta/2}^\alpha$ . Reasoning conditionally on  $X$ , we recognize in  $\frac{1}{n(n-1)} \sum_{i \neq j}^n K(X_i, X_j) \epsilon_i \epsilon_j := T_K^{(0)}$  be a Gaussian chaos, as defined by [De la Pena and Giné \(2012\)](#), of the form  $Z = \sum_{i \neq i'} x_{i,i'} \epsilon_i \epsilon_{i'}$ , where  $x_{i,i'}$ 's are some real deterministic numbers and  $(\epsilon_i)_i$  is a sequence of i.i.d Gaussian variables. Corollary 3.26 of [De la Pena and Giné \(2012\)](#) states that there exists some absolute constant  $\kappa > 0$  such that if  $\gamma^2 = \mathbb{E}[Z^2] = \sum_{i \neq i'} x_{i,i'}^2$ . Then

$$\mathbb{E} \left[ \exp \left( \frac{|Z|}{\kappa \gamma} \right) \right] \leq 2.$$

Hence by Markov's inequality,

$$\mathbb{P} \left( |Z| \geq \kappa \gamma \ln \left( \frac{2}{\alpha} \right) \right) \leq \alpha. \quad (3.6.9)$$

Applying the result (3.6.9) for  $T_K^{(0)}$  with

$$\gamma^2 = \frac{\sigma^4}{n^2(n-1)^2} \sum_{i \neq j=1}^n K_{ij}^2, \quad (3.6.10)$$

we have

$$\mathbb{P}_{(H_0)} \left( T_K \geq \frac{\kappa \sigma^2}{n(n-1)} \ln \left( \frac{4}{\alpha} \right) \sqrt{\sum_{i \neq j=1}^n K_{ij}^2} \right) \leq \frac{\alpha}{2}. \quad (3.6.11)$$

On the other hand, we have, under  $H_0$ ,  $\hat{\sigma}_n^2 = \hat{\sigma}_{n,\epsilon}^2$  where

$$\hat{\sigma}_{n,\epsilon}^2 = \frac{\sigma^2}{n} \sum_{i=1}^{n/2} (\epsilon'_{2i-1} - \epsilon'_{2i})^2 = \frac{2\sigma^2}{n} \sum_{i=1}^{n/2} \left( \frac{\epsilon'_{2i-1} - \epsilon'_{2i}}{\sqrt{2}} \right)^2 \sim \frac{2\sigma^2}{n} \chi_{(n/2)}.$$

Since the variables  $Z_i = \epsilon'_{2i-1} - \epsilon'_{2i}$ ,  $i = 1, \dots, n/2$  are i.i.d standard Gaussian variables. Using the Lemma 8.1 in [Birgé \(2001\)](#), we obtain

$$\forall x > 0, \mathbb{P}_{(H_0)} \left( \frac{2\sigma^2}{n} \chi_{(n/2)} \leq \sigma^2 - \frac{4\sigma^2}{\sqrt{2n}} \sqrt{x} \right) \leq e^{-x}. \quad (3.6.12)$$

Choosing  $x = \ln(2/\alpha)$ , [\(3.6.12\)](#) leads to

$$\mathbb{P}_{(H_0)} \left( \hat{\sigma}_n^2 \leq \sigma^2 - \frac{4\sigma^2}{\sqrt{2n}} \sqrt{\ln \left( \frac{2}{\alpha} \right)} \right) \leq \frac{\alpha}{2}. \quad (3.6.13)$$

Moreover, we have

$$\begin{aligned} \mathbb{P}_{(H_0)} \left( \frac{T_K}{\hat{\sigma}_n^2} \geq \frac{\frac{\kappa\sigma^2}{n(n-1)} \ln \left( \frac{2}{\alpha} \right) \sqrt{\sum_{i \neq j=1}^n K_{ij}^2}}{\sigma^2 \left( 1 - \frac{2\sqrt{2}}{\sqrt{n}} \sqrt{\ln \left( \frac{2}{\alpha} \right)} \right)} \right) &\leq \mathbb{P}_{(H_0)} \left( T_K \geq \frac{\kappa\sigma^2}{n(n-1)} \ln \left( \frac{2}{\alpha} \right) \sqrt{\sum_{i \neq j=1}^n K_{ij}^2} \right) \\ &+ \mathbb{P}_{(H_0)} \left( \frac{1}{\hat{\sigma}_n^2} \geq \frac{1}{\sigma^2 \left( 1 - \frac{2\sqrt{2}}{\sqrt{n}} \sqrt{\ln \left( \frac{2}{\alpha} \right)} \right)} \right). \end{aligned}$$

From [\(3.6.11\)](#) and [\(3.6.13\)](#), we obtain

$$\mathbb{P}_{(H_0)} \left( \frac{T_K}{\hat{\sigma}_n^2} \geq \frac{\frac{\kappa\sigma^2}{n(n-1)} \ln \left( \frac{2}{\alpha} \right) \sqrt{\sum_{i \neq j=1}^n K_{ij}^2}}{\sigma^2 \left( 1 - \frac{2\sqrt{2}}{\sqrt{n}} \sqrt{\ln \left( \frac{2}{\alpha} \right)} \right)} \right) \leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha. \quad (3.6.14)$$

This implies

$$\mathbb{P}_{(H_0)} \left( V_K \geq \frac{\frac{\kappa\sigma^2}{n(n-1)} \ln \left( \frac{2}{\alpha} \right) \sqrt{\sum_{i \neq j=1}^n K_{ij}^2}}{\sigma^2 \left( 1 - \frac{2\sqrt{2}}{\sqrt{n}} \sqrt{\ln \left( \frac{2}{\alpha} \right)} \right)} \right) \leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.$$

Thus the  $(1 - \alpha)$  quantile of  $V_K^{(0)}$  conditionally on  $X$  satisfies

$$q_{K,1-\alpha}^{(X)} \leq \frac{\frac{\kappa\sigma^2}{n(n-1)} \ln \left( \frac{2}{\alpha} \right) \sqrt{\sum_{i \neq j=1}^n K_{ij}^2}}{\sigma^2 \left( 1 - \frac{2\sqrt{2}}{\sqrt{n}} \sqrt{\ln \left( \frac{2}{\alpha} \right)} \right)}. \quad (3.6.15)$$

Taking  $n \geq 32 \ln \left( \frac{2}{\alpha} \right)$ , so  $\sqrt{n} \geq 4\sqrt{2} \sqrt{\ln \left( \frac{2}{\alpha} \right)}$ , [\(3.6.15\)](#) returns to

$$q_{K,1-\alpha}^{(X)} \leq \frac{2\kappa}{n(n-1)} \ln \left( \frac{2}{\alpha} \right) \sqrt{\sum_{i \neq j=1}^n K_{ij}^2}. \quad (3.6.16)$$

Hence  $q_{K,1-\beta/2}^\alpha$  is upper bounded by the  $(1 - \beta/2)$  quantile of  $\frac{2\kappa}{\sqrt{n(n-1)}} \ln\left(\frac{2}{\alpha}\right) \sqrt{\frac{1}{n(n-1)} \sum_{i \neq j=1}^n K_{ij}^2}$ .

We define

$$U_n = \frac{1}{n(n-1)} \sum_{i \neq j=1}^n K_{ij}^2.$$

We use Markov's inequality again for the nonnegative random variable  $U_n$ , we obtain for any  $\delta > 0$

$$\mathbb{P}_f(U_n > \delta) \leq \frac{\mathbb{E}(U_n)}{\delta}. \quad (3.6.17)$$

We have

$$\mathbb{E}[U_n] = \int_{E^2} K^2(x, y) d\nu(x) d\nu(y). \quad (3.6.18)$$

Choosing  $\delta = 2 \int_{E^2} K^2(x, y) d\nu(x) d\nu(y) / \beta$ , (3.6.17) returns to

$$\mathbb{P}_f\left(U_n > \frac{2 \int_{E^2} K^2(x, y) d\nu(x) d\nu(y)}{\beta}\right) \leq \frac{\beta}{2},$$

and

$$q_{K,1-\beta/2}^\alpha \leq \frac{2\kappa}{\sqrt{n(n-1)}} \ln\left(\frac{2}{\alpha}\right) \sqrt{\frac{2 \int_{E^2} K^2(x, y) d\nu(x) d\nu(y)}{\beta}},$$

which concludes the proof.

### 3.6.2 Proof of Theorem 3.2

For all symmetric kernel function  $K$ , we have

$$\mathbb{E}(T_K) = \langle K[f], f \rangle = \frac{1}{2} (\|f\|^2 + \|K[f]\|^2 - \|f - K[f]\|^2). \quad (3.6.19)$$

On the other hand

$$A_K \leq \frac{(\|f\|_\infty^2 + \sigma^2) \|K[f]\|^2}{n}.$$

Let  $C_K$  be an upper bound for  $\int_{E^2} K^2(x, y) d\nu(x) d\nu(y)$ , we have

$$B_K \leq \frac{(\|f\|_\infty^2 + \sigma^2)^2 C_K}{n(n-1)}$$

From Proposition 3.1, the bounds for  $A_K$  and  $B_K$  and the inequality  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$  for all  $a \geq 0, b \geq 0$ , we deduce that  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$  as soon as,

$$\begin{aligned} \|f\|^2 + \|K[f]\|^2 - \|f - K[f]\|^2 &\geq 4\sqrt{2}\kappa D_{n,\beta} \ln\left(\frac{2}{\alpha}\right) \sqrt{\frac{C_K}{n(n-1)\beta}} + 4\sqrt{2} (\|f\|_\infty^2 + \sigma^2) \sqrt{\frac{C_K}{n(n-1)\beta}} \\ &\quad + 8 \|K[f]\| \sqrt{\frac{(\|f\|_\infty^2 + \sigma^2)}{n\beta}}. \end{aligned}$$

By using the elementary inequality  $2cd \leq c^2 + d^2$  with  $c = \|K[f]\|$  and  $d = 4\sqrt{\frac{(\|f\|_\infty^2 + \sigma^2)}{n\beta}}$  in the right hand side of the above condition, the above condition holds if

$$\begin{aligned} \|f\|^2 &\geq \|f - K[f]\|^2 + \frac{16(\|f\|_\infty^2 + \sigma^2)}{n\beta} \\ &\quad + \frac{4\sqrt{C_K}}{\sqrt{n(n-1)\beta}} \left( \kappa D_{n,\beta} \ln\left(\frac{2}{\alpha}\right) + \sqrt{2}(\|f\|_\infty^2 + \sigma^2) \right). \end{aligned}$$

### 3.6.3 Proof of Corollary 3.5 and 3.7.

Under the hypothesis of corollary 3.5,

$$K(x, y) = \sum_{\lambda \in \Lambda} \phi_\lambda(x) \phi_\lambda(y),$$

and the linear space  $S$  generated by the functions  $(\phi_\lambda, \lambda \in \Lambda)$  is of dimension  $D$ . Hence, we have

$$\int_{[0,1]^2} \left( \sum_{\lambda \in \Lambda} \phi_\lambda(x) \phi_\lambda(y) \right)^2 d\nu(x) d\nu(y) \leq D.$$

Thus, we can take  $C_K = D$ .

Second, under choice of the Gaussian kernel defined by (3.2.6), we recall that

$$K(x, y) = \frac{1}{h} k\left(\frac{x-y}{h}\right), \quad \text{for } (x, y) \in \mathbb{R}^2 \quad (3.6.20)$$

where  $k(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$ , for all  $u \in \mathbb{R}$  and  $h$  is a positive bandwidth.

We have

$$\begin{aligned} \int_{\mathbb{R}^2} K^2(x, y) \nu(x) \nu(y) d(x) d(y) &= \frac{1}{2\pi h^2} \int_{\mathbb{R}^2} e^{-\frac{(x-y)^2}{2h^2}} \nu(x) \nu(y) d(x) d(y) \\ &\leq \frac{1}{2\pi h^2} \int_{\mathbb{R}^2} e^{-\frac{u^2}{2}} h du \nu(x) \nu(x - uh) dx \\ &\leq \frac{\|\nu\|_\infty}{h\sqrt{2\pi}} \left( \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}} du \right) \left( \int_{\mathbb{R}} \nu(x) dx \right) \\ &\leq \frac{\|\nu\|_\infty}{h\sqrt{2\pi}}. \end{aligned}$$

Hence, we can choose  $C_K = \frac{\|\nu\|_\infty}{h\sqrt{2\pi}}$ .

### 3.6.4 Proof of Proposition 3.6

For all  $J \geq 0$ , we set  $D = 2^J$  be the dimension of  $S_J$ . Let us assume that  $f \in \mathcal{B}_{2,\infty}^\delta(\mathbb{R})$ , it implies

$$\|f - \Pi_{S_J}(f)\|^2 \leq R^2 2^{-2J\delta}. \quad (3.6.21)$$

We obtain from Corollary 3.5 that there exists

$$C(\alpha, \beta, \sigma, \|f\|_\infty) > 0,$$

such that  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$  if

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, \|f\|_\infty) \left[ 2^{-2J\delta} + \frac{2^{J/2}}{n} \right]. \quad (3.6.22)$$

In this case, we see that the right hand side of (3.3.5) reproduces a bias-variance decomposition close to the bias-variance decomposition for projection estimators, with the bias term  $R^2 2^{-2J\delta}$  and the variance term  $2^{J/2}/n$ . The optimal choice of  $J$  satisfies

$$R^2 2^{-2J\delta} = \frac{2^{J/2}}{n}.$$

Thus, we obtain the optimal choice  $J^*$ ,

$$J^* = \left\lceil \log_2 \left( n^{2/(1+4\delta)} \right) \right\rceil,$$

leading to the desired result.

### 3.6.5 Proof of Proposition 3.8

Considering (3.3.6), we mainly have to find a sharp upper bound for  $\|f - k_h * f\|^2$  when  $f \in \mathcal{S}_\delta(R)$ . Plancherel's theorem gives that when  $f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ ,

$$\begin{aligned} (2\pi)\|f - k_h * f\|^2 &= \left\| (1 - \widehat{k}_h) (\widehat{f}) \right\|^2 \\ &= \int_{\mathbb{R}} \left| 1 - \widehat{k}(2^{-l}u) \right|^2 (u) \widehat{f}^2(u) du. \end{aligned}$$

We assume that  $\|\widehat{k}\|_\infty < +\infty$  and

$$\text{Ess sup}_{u \in \mathbb{R} \setminus \{0\}} \frac{|1 - \widehat{k}(u)|}{|u|} \leq C,$$

for some  $C > 0$ . There also exists some constant  $C(\delta) > 0$  such that

$$\text{Ess sup}_{u \in \mathbb{R} \setminus \{0\}} \frac{|1 - \widehat{k}(u)|}{|u|^\delta} \leq C(\delta).$$

Then

$$\|f - k_h * f\|^2 \leq \frac{C(\delta)}{2\pi} \int_{\mathbb{R}} |2^{-l}u|^{2\delta} \widehat{f}^2(u) du$$

and since  $f \in \mathcal{S}_\delta(R)$ ,

$$\|f - k_h * f\|^2 \leq 2^{-2\delta l} C(\delta) R^2.$$

We obtain from corollary 3.7 that there exists

$$C(\alpha, \beta, \sigma, R, \|f\|_\infty) > 0,$$

such that  $\mathbb{P}_f(\Phi_{K,\alpha} = 0) \leq \beta$  if

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, \|f\|_\infty) \left[ 2^{-2\delta l} + \frac{2^{l/2}}{n} \right]. \quad (3.6.23)$$

In this case, we see that the right hand side of (3.6.23) reproduces a bias-variance decomposition with the bias term  $2^{-2\delta l}$  and the variance term  $2^{l/2}/n$ . The optimal choice of  $l$  satisfies

$$2^{-2\delta l} = \frac{2^{l/2}}{n}.$$

Thus, we obtain the optimal choice  $l^*$  as follows.

$$l^* = \left\lceil \log_2 \left( n^{2/(1+4\delta)} \right) \right\rceil,$$

leading to the desired result.

### 3.6.6 Proof of Theorem 3.9, Corollary 3.10 and 3.12.

We have

$$\begin{aligned} \mathbb{P}_{(H_0)}(\Phi_\alpha = 1) &= \mathbb{P} \left[ \sup_{m \in \mathcal{M}} \left( V_{K_m} - q_{m,1-u_\alpha^{(X)}}^{(X)} e^{-w_m} \right) > 0 \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \mathbb{1} \left\{ \sup_{m \in \mathcal{M}} \left( V_{K_m} - q_{m,1-u_\alpha^{(X)}}^{(X)} e^{-w_m} \right) > 0 \right\} \middle| X \right] \right]. \end{aligned}$$

We have,

$$\mathbb{E} \left[ \mathbb{1} \left\{ \sup_{m \in \mathcal{M}} \left( V_{K_m} - q_{m,1-u_\alpha^{(X)}}^{(X)} e^{-w_m} \right) > 0 \right\} \middle| X \right] \leq \alpha,$$

by definition of  $u_\alpha^{(X)}$ , which implies that  $\mathbb{P}_{(H_0)}(\Phi_\alpha = 1) \leq \alpha$ .

On the other hand, we know that  $u_\alpha^{(X)} \geq \alpha$ . Setting  $\alpha_m = \alpha e^{-w_m}$ , we have

$$\begin{aligned} \mathbb{P} \left( \exists m \in \mathcal{M}, V_{K_m} > q_{K_m,1-u_\alpha^{(X)}}^{(X)} e^{-w_m} \right) &\geq \mathbb{P} \left( \exists m \in \mathcal{M}, V_{K_m} > q_{K_m,1-\alpha_m}^{(X)} \right) \\ &\geq 1 - \mathbb{P} \left( \forall m \in \mathcal{M}, V_{K_m} \leq q_{K_m,1-\alpha_m}^{(X)} \right) \geq 1 - \inf_{m \in \mathcal{M}} \mathbb{P} \left( V_{K_m} \leq q_{K_m,1-\alpha_m}^{(X)} \right) \\ &\geq 1 - \beta, \end{aligned}$$

as soon as there exists  $m$  in  $\mathcal{M}$  such that

$$\mathbb{P} \left( V_{K_m} \leq q_{K_m,1-\alpha_m}^{(X)} \right) \leq \beta.$$

We can now apply Corollary 3.5 and 3.7 with  $\alpha_m = \alpha e^{-w_m}$ , so we replace  $\ln(2/\alpha)$  by  $(\ln(2/\alpha) + w_m)$  for desired results in Corollary 3.10 and 3.12.

### 3.6.7 Proof of Corollary 3.11.

Considering (3.4.6), we aim to find a sharp upper bound for the right hand side of the inequality when  $f \in \mathcal{B}_{2,\infty}^\delta(R, R')$ . Let us assume that  $f \in \mathcal{B}_{2,\infty}^\delta(R, R')$ . Then  $f \in \mathcal{B}_{2,\infty}^\delta(R)$ , we have

$$\|f - \Pi_{S_J}(f)\|^2 \leq C(\delta)R^2 2^{-2\delta J},$$

and

$$C(\alpha, \beta, \sigma, R, \|f\|_\infty) \leq C(\alpha, \beta, \sigma, R, R').$$

Hence (3.4.6) can be upper bounded by

$$C(\alpha, \beta, \sigma, R, R') \inf_{J \in \mathcal{M}_{\bar{J}}} \left\{ 2^{-2\delta J} + \ln(2 + J) \frac{2^{J/2}}{n} \right\}.$$

Taking

$$J^{**} = \left\lceil \log_2 \left( \left( \frac{n}{\ln \ln(n)} \right)^{\frac{2}{4\delta+1}} \right) \right\rceil,$$

$$\begin{aligned} C(\alpha, \beta, \sigma, R, R') & \inf_{J \in \mathcal{M}_{\bar{J}}} \left\{ 2^{-2\delta J} + \ln(2 + J) \frac{2^{J/2}}{n} \right\} \\ & \leq C(\alpha, \beta, \sigma, R, R') \left\{ 2^{-2\delta J^{**}} + \ln(2 + J^{**}) \frac{2^{J^{**}/2}}{n} \right\} \\ & \leq C(\alpha, \beta, \sigma, R, R') \left( \frac{n}{\ln \ln(n)} \right)^{-\frac{4\delta}{4\delta+1}}. \end{aligned}$$

That leads to  $\mathbb{P}_f(\Phi_\alpha^{(1)} = 0) \leq \beta$  if

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, R') \left( \frac{\ln \ln(n)}{n} \right)^{\frac{4\delta}{4\delta+1}}.$$

### 3.6.8 Proof of Corollary 3.12.

Considering (3.4.10), we aim to find a sharp upper bound for the right hand side of the inequality when  $f \in \mathcal{S}_\delta(R, R')$ . Let us assume that  $f \in \mathcal{S}_\delta(R, R')$ . Similarly, with regards to the proof of Proposition 3.8, we have

$$\|f - k_l * f\|^2 \leq 2^{-2\delta l} C(\delta) R^2,$$

and

$$C(\alpha, \beta, \sigma, \|f\|_\infty) \leq C(\alpha, \beta, \sigma, R').$$

Hence (3.4.10) can be upper bounded by

$$C(\alpha, \beta, \sigma, R, R') \inf_{l \in \mathcal{L}} \left\{ 2^{-2\delta l} + \frac{w_l}{n\sqrt{2^{-l}}} \right\}.$$

Choosing

$$l^{**} = \left\lceil \log_2 \left( \left( \frac{n}{\ln \ln(n)} \right)^{\frac{2}{1+4\delta}} \right) \right\rceil,$$

$$\begin{aligned} C(\alpha, \beta, \sigma, R, R') \inf_{l \in \mathcal{L}} \left\{ 2^{-2\delta l} + \frac{w_l}{n\sqrt{2^{-l}}} \right\} \\ \leq C(\alpha, \beta, \sigma, R, R') \left\{ 2^{-2\delta l^{**}} + \frac{w_{l^{**}}}{n\sqrt{2^{-l^{**}}}} \right\} \\ \leq C(\alpha, \beta, \sigma, R, R') \left( \frac{n}{\ln \ln(n)} \right)^{-\frac{4\delta}{4\delta+1}}. \end{aligned}$$

That leads to  $\mathbb{P}_f(\Phi_\alpha^{(1)} = 0) \leq \beta$  if

$$\|f\|^2 \geq C(\alpha, \beta, \sigma, R, R') \left( \frac{\ln \ln(n)}{n} \right)^{\frac{4\delta}{4\delta+1}}.$$

## Acknowledgement

I gratefully thank to Professor Béatrice Laurent of Institut National des Sciences Appliquées de Toulouse and Professor Jean-Michel Loubes of Institut de Mathématiques de Toulouse for supporting me in the best ideas and comments.





# Chapitre 4

## Conclusion et perspectives

This work has proposed two mathematical frameworks to deal with interesting information conveyed by weak signals. Firstly, a new estimated function has been introduced for the regression model with distribution inputs. More precisely, we have effectively used a class of positive definite kernels produced by Wasserstein distance. The building in [Bachoc et al. \(2017\)](#) is a such kind of the universal kernel. Researching the universal kernel theories, we have indicated a very good property of our universal kernel to build a RKHS. We have then obtained a particular estimation from Representer theorem for our distribution regression problem: these two pieces of works have shown that the relation between the random distribution and the real number response can be learnt by directly using the regularized empirical risk over RKHS. More interestingly, we have studied the TEOAE curve of each individual in the human population as a distribution well. We have then investigated the potential relationship existing between the age of an individual and its TEOAE response. We have shown that the TEOAE response is clearly a function of the age, and the effect of age in hearing issues is probably deeply related to the change of cochlear. This is a new interesting approach in the field of Biostatistics, in which we have indicated the evolution of hearing capacity under the statistical domain of a distribution regression model. Moreover, we believe that our paper tackles an important issue for data science experts willing to predict problems in regression with probability distributions as an input.

### **Hearing evolution of non-human primate species**

The relationship between the age and hearing sensitivity of non-human primates is worth further study in the future. For example, Figure 4.1 shows that the real age (blue points) and predicted age (red points) of *Macaca sylvanus* (also known as Rhesus Macaque) calculated by our regression estimation, appears to be less precise than for humans, despite their statistical significance. The much more variable response in macaque compare to humans might be the reason for such a difference between the human and macaque, but the lower number of individuals can also bias the prediction. More analyses with newly collected data are needed, but what it is

remarkable is that, despite this not being the first time such kind of data has been collected and analyzed for macaque, our model seems to predict the age according to the TEOAE response. For future research, we can better improve the mathematical model in this paper to be closer to the biological data, particularly, the TEOAE response of non-human primates.

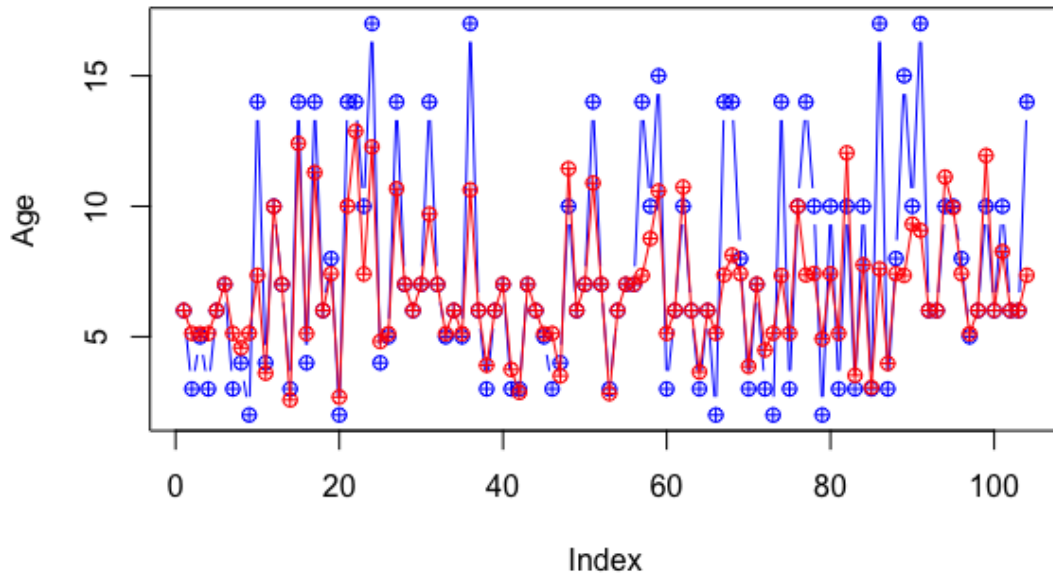


Figure 4.1: Real and Predicted Age of *Macaca Sylvanus* by applying our framework.

## Multiple distribution regression model with a Reproducing Kernel Hilbert Space approach

We have studied about the regression model with one-dimensional distribution entries under a RKHS approach in Part 2. The extension of this work on distributions for general dimensions should be addressed in a further work. In the general dimension case, in order to build a positive definite kernel from the Wasserstein distance, [Bachoc et al. \(2018\)](#) have associated to each input distribution its optimal transport map to a reference distribution. The authors have then provided positive definite kernels on the Hilbert space corresponding to these optimal transport maps. However, we do not know any estimation for optimal transport maps. Recently, [Hütter and Rigollet \(2019\)](#) have established an estimation of the transport map which is up to logarithmic factors of the minimax rate. These are the first minimax estimation rates for transport maps in the general dimension. Thus it is appreciated to

make a link between these studies in future in the domain of regression models with general dimensional distribution entries.

Secondly, the aggregated kernel based testing procedures has been constructed for testing the nullity of the regression function. It is available when the variance of the observations is unknown and does not depend on any prior information on the alternative. Our test statistics have been based on a single kernel function which can be chosen either as a projection or Gaussian kernel and we have proposed an estimation for the unknown variance. Our tests are exactly (and not only asymptotically) of level  $\alpha$ . We have obtained the optimal non-asymptotic conditions on the alternative which guarantee that the probability of second kind error is at most equal to a prescribed level. However, the testing procedure that we have introduced hereafter also intended to overcome the question of calibrating the choice of kernel and/or the parameters of the kernel. To avoid these problems, the aggregated testing procedures have been constructed to be of level  $\alpha$  and the loss in second kind error due to the aggregation, when unavoidable, is as small as possible. The multiple tests satisfy non-asymptotic properties and are adaptive in the minimax sense over several classes of regular alternatives.

## Testing the variability of TEOAEs between males and females

Basing on Part 3, we would be very interested in testing the difference existing between males and females in the human population but also in non-human primates. We will consider, the following models,

$$\begin{aligned} Y_i^{(1)} &= f^{(1)}(\mu_i) + \sigma\epsilon_i^{(1)}, \\ Y_i^{(2)} &= f^{(2)}(\mu_i) + \sigma\epsilon_i^{(2)}, \end{aligned}$$

where  $\mu_i$  is considered as probability distributions corresponding to TEOAE curves,  $Y^{(j)}$ ,  $j = 1, 2$  instead considered by functional objects, they are considered by classified objects, for instance, males and females.

We expect to test the hypothesis,

$$(H_0) : f^{(1)} - f^{(2)} = 0,$$

with the alternative

$$(H_1) : f^{(1)} - f^{(2)} \neq 0.$$

Resolving this problem could be interesting question for a future project.

## Big data analytics

This research investigates analyses of TEOAE profiles from different populations (Asia, Europe, Africa) with particular interest in investigating human and non-human primates.



# Bibliography

- D Aidan, P Lestang, P Avan And, and P Bonfils. Characteristics of transient-evoked otoacoustic emissions (teoes) in neonates. *Acta oto-laryngologica*, 117(1):25–30, 1997a.
- Darmody Aidan, Paul De Lestang, Paul Avan, and Pierre Bonfils. Characteristics of transient-evoked otoacoustic emissions (teoes) in neonates. *Acta oto-laryngologica*, 117 1:25–30, 1997b.
- Pedro C Alvarez-Esteban, E del Barrio, JA Cuesta-Albertos, and C Matrán. Wide consensus for parallelized inference. *arXiv preprint arXiv:1511.05350*, 2015.
- Jean-Marc Azaïs. Le modèle linéaire par l'exemple. 2006.
- François Bachoc, Fabrice Gamboa, Jean-Michel Loubes, and Nil Venet. A gaussian process regression model for distribution inputs. *IEEE Transactions on Information Theory*, 2017.
- Francois Bachoc, Alexandra Suvorikova, Jean-Michel Loubes, and Vladimir Spokoiny. Gaussian process forecast with multidimensional distributional entries. *arXiv preprint arXiv:1805.00753*, 2018.
- Yannick Baraud, Sylvie Huet, Béatrice Laurent, et al. Adaptive tests of linear hypotheses by model selection. *The Annals of Statistics*, 31(1):225–251, 2003.
- Yannick Baraud et al. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.
- Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
- Lucien Birgé. An alternative point of view on lepski's method. *Lecture Notes-Monograph Series*, pages 113–133, 2001.
- Pierre Bonfils, Yves Bertrand, and Alain Uziel. Evoked otoacoustic emissions: normative data and presbycusis. *Audiology*, 27(1):27–35, 1988.
- Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- Thi Thien Trang Bui. Aggregated kernel based tests for signal detection in a regression model. *arXiv preprint arXiv:1904.02965*, 2019.
- Thi Thien Trang Bui, JM Loubes, et al. Distribution regression model with a reproducing kernel hilbert space approach. *arXiv preprint arXiv:1806.10493*, 2018.
- Cristina Butucea and Karine Tribouley. Nonparametric homogeneity tests. *Journal of statistical planning and inference*, 136(3):597–639, 2006.
- Jane W Cassidy and Karen M Ditty. Gender differences among newborns on a transient otoacoustic emissions test for hearing. *Journal of Music Therapy*, 38(1):28–35, 2001.

- I Castillo, C Lévy-Leduc, and C Matias. Exact adaptive estimation of the shape of a periodic function with unknown period corrupted by white noise. *Mathematical methods of statistics*, 15(2):146–175, 2006.
- Jeng-Min Chiou, Hans-Georg Müller, and Jane-Ling Wang. Functional response models. *Statistica Sinica*, pages 675–693, 2004.
- Andreas Christmann and Ingo Steinwart. Universal kernels on non-standard input spaces. In *Advances in neural information processing systems*, pages 406–414, 2010.
- David Y Chung, Keith Mason, R Patrick Gannon, and Glenn N Willson. The ear effect as a function of age and hearing loss. *The Journal of the Acoustical Society of America*, 73(4):1277–1282, 1983.
- Lionel Collet, Annie Moulin, Myriam Gartner, and Alain Morgon. Age-related changes in evoked otoacoustic emissions. *Annals of Otolaryngology, Rhinology & Laryngology*, 99(12):993–997, 1990.
- Lionel Collet, Myriam Gartner, Evelyne Veuillet, Annie Moulin, and Alain Morgon. Evoked and spontaneous otoacoustic emissions. *Brain and Development*, 15(4):249–252, jul 1993. doi: 10.1016/0387-7604(93)90018-4. URL <https://doi.org/10.1016%2F0387-7604%2893%2990018-4>.
- Michael G Cowling. Harmonic analysis on semigroups. *Annals of Mathematics*, pages 267–283, 1983.
- Peter Dallos. The role of outer hair cells in cochlear function. In *Progress in clinical and biological research*. Alan R. Liss, 1985.
- Victor De la Pena and Evarist Giné. *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012.
- Miguel A Delgado. Testing the equality of nonparametric regression curves. 1992.
- Evgenia Dimitriadou, Kurt Hornik, Friedrich Leisch, David Meyer, Andreas Weingessel, and Maintainer Friedrich Leisch. Package ‘e1071’. *R Software package, available at <http://cran.rproject.org/web/packages/e1071/index.html>*, 2009.
- Alessandra Spada Durante and Renata Mota Mamede Carvalho. Changes in transient evoked otoacoustic emissions contralateral suppression in infants. *Pró-Fono Revista de Atualização Científica*, 18(1):49–56, 2006.
- Jos J Eggermont. *Hearing Loss: Causes, Prevention, and Treatment*. Academic Press, 2017.
- Paul Embrechts and Marius Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013.
- RL Eubank and VN LaRiccia. Testing for no effect in nonparametric regression. *Journal of statistical planning and inference*, 36(1):1–14, 1993.
- Cynthia G Fowler, Kirstin Beach Chiasson, Tami Hanson Leslie, Denise Thomas, T Mark Beasley, Joseph W Kemnitz, and Richard Weindruch. Auditory function in rhesus monkeys: effects of aging and caloric restriction in the wisconsin monkeys five years later. *Hearing research*, 261(1-2):75–81, 2010.
- Magalie Fromont and Céline Lévy-Leduc. Adaptive tests for periodic signal detection with applications to laser vibrometry. *ESAIM: Probability and Statistics*, 10:46–75, 2006.

- Magalie Fromont, Béatrice Laurent, Patricia Reynaud-Bouret, et al. Adaptive tests of homogeneity for a poisson process. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 47, pages 176–213. Institut Henri Poincaré, 2011.
- Magalie Fromont, Matthieu Lerasle, Patricia Reynaud-Bouret, et al. Kernels based tests with non-asymptotic bootstrap approaches for two-sample problems. In *Conference on Learning Theory*, pages 23–1, 2012.
- Magalie Fromont, Béatrice Laurent, Patricia Reynaud-Bouret, et al. The two-sample problem for poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *The Annals of Statistics*, 41(3):1431–1461, 2013.
- Thomas Gold. Hearing. ii. the physical basis of the action of the cochlea. *Proceedings of the Royal Society of London B: Biological Sciences*, 135(881):492–498, 1948.
- Thibaut Le Gouic and Jean-Michel Loubes. Existence and consistency of wasserstein barycenters. *Probability Theory and Related Fields*, 168(3-4):901–917, aug 2016. doi: 10.1007/s00440-016-0727-z. URL <https://doi.org/10.1007%2Fs00440-016-0727-z>.
- Arthur Gretton, Karsten M Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex J Smola. A kernel method for the two-sample-problem. In *Advances in neural information processing systems*, pages 513–520, 2007.
- Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(Mar):723–773, 2012.
- James Wilbur Hall. *Handbook of otoacoustic emissions*. Cengage Learning, 2000.
- Peter Hall and Jeffrey D Hart. Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, 85(412):1039–1049, 1990.
- Wolfgang Hardle and J Steve Marron. Semiparametric comparison of regression curves. *The Annals of Statistics*, pages 63–89, 1990.
- AJ Hudspeth. The cellular basis of hearing: the biophysics of hair cells. *Science*, 230(4727):745–752, 1985a.
- AJ Hudspeth. Models for mechano-electrical transduction by hair cells. *Progress in clinical and biological research*, 176:193–205, 1985b.
- Marie Huskova and Paul Janssen. Consistency of the generalized bootstrap for degenerate u-statistics. *The Annals of Statistics*, pages 1811–1823, 1993.
- Jan-Christian Hütter and Philippe Rigollet. Minimax rates of estimation for smooth optimal transport maps. *arXiv preprint arXiv:1905.05828*, 2019.
- Yuri I Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. i, ii, iii. *Math. Methods Statist*, 2(2):85–114, 1993.
- NJ Johnsen and C Elberling. Evoked acoustic emissions from the human ear: I. equipment and response parameters. *Scandinavian audiology*, 11(1):3–12, 1982.
- NJ Johnsen, P Bagi, and C Elberling. Evoked acoustic emissions from the human ear: Iii. findings in neonates. *Scandinavian audiology*, 12(1):17–24, 1983.
- NJ Johnsen, P Bagi, J Parbo, and C Elberling. Evoked acoustic emissions from the human ear iv. final results in 100 neonates. *Scandinavian Audiology*, 17(1):27–34, 1988.



- Philip X Joris, Christopher Bergevin, Radha Kalluri, Myles Mc Laughlin, Pascal Michelet, Marcel van der Heijden, and Christopher A Shera. Frequency selectivity in old-world monkeys corroborates sharp cochlear tuning in humans. *Proceedings of the National Academy of Sciences*, 108(42):17516–17520, 2011.
- Hachem Kadri, Emmanuel Duflos, Philippe Preux, Stéphane Canu, and Manuel Davy. Nonlinear functional regression: a functional rkhs approach. In *Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS'10)*, volume 9, pages 374–380, 2010.
- Douglas H Keefe, Michael P Gorga, Walt Jesteadt, and Lynette M Smith. Ear asymmetries in middle-ear, cochlear, and brainstem responses in human infants. *The Journal of the Acoustical Society of America*, 123(3):1504–1512, 2008.
- David T Kemp. Stimulated acoustic emissions from within the human auditory system. *The Journal of the Acoustical Society of America*, 64(5):1386–1391, 1978.
- David T Kemp. Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea. *Archives of oto-rhino-laryngology*, 224(1-2):37–45, 1979.
- David T Kemp and Siobhan Ryan. Otoacoustic emission tests in neonatal screening programmes. *Acta Oto-Laryngologica*, 111(sup482):73–84, 1991.
- DT Kemp. Towards a model for the origin of cochlear echoes. *Hearing Research*, 2(3-4):533–548, 1980.
- DT Kemp. Physiologically active cochlear micromechanics-one source of tinnitus. *Tinnitus*, pages 54–81, 1981.
- George Kimeldorf and Grace Wahba. Some results on tchebycheffian spline functions. *Journal of mathematical analysis and applications*, 33(1):82–95, 1971.
- Eileen King, Jeffrey D Hart, and Thomas E Wehrly. Testing the equality of two regression curves using linear smoothers. *Statistics & Probability Letters*, 12(3):239–247, 1991.
- Eileen Catherine King. *A test for the equality of two regression curves based on kernel smoothers*. PhD thesis, Texas A & M University, 1988.
- Soheil Kolouri, Yang Zou, and Gustavo K. Rohde. Sliced Wasserstein kernels for probability distributions. *CoRR*, abs/1511.03198, 2015. URL <http://arxiv.org/abs/1511.03198>.
- Soheil Kolouri, Gustavo K Rohde, and Heiko Hoffmann. Sliced wasserstein distance for learning gaussian mixture models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3427–3436, 2018.
- Kaori Kon, Masumi Inagaki, and Makiko Kaga. Developmental changes of distortion product and transient evoked otoacoustic emissions in different age groups. *Brain and Development*, 22(1):41–46, jan 2000. doi: 10.1016/s0387-7604(99)00114-x. URL [https://doi.org/10.1016/S0387-7604\(99\)00114-x](https://doi.org/10.1016/S0387-7604(99)00114-x).
- Michael H Kutner, Chris Nachtsheim, and John Neter. *Applied linear regression models*. McGraw-Hill/Irwin, 2004.
- Thibaut Le Gouic and Jean-Michel Loubes. Existence and consistency of wasserstein barycenters. *Probability Theory and Related Fields*, 168(3-4):901–917, 2017.
- Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

- Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3): 18–22, 2002.
- Jinfeng Liu and Ningyu Wang. Effect of age on click-evoked otoacoustic emission: A systematic review. *Neural regeneration research*, 7(11):853, 2012.
- Alex B McBratney, ML Mendonça Santos, and Budiman Minasny. On digital soil mapping. *Geoderma*, 117(1-2):3–52, 2003.
- Dennis McFadden. A speculation about the parallel ear asymmetries and sex differences in hearing sensitivity and otoacoustic emissions. *Hearing research*, 68(2):143–151, 1993.
- Dennis McFadden. Sex differences in the auditory system. *Developmental Neuropsychology*, 14(2-3):261–298, 1998. doi: 10.1080/87565649809540712. URL <https://doi.org/10.1080/87565649809540712>.
- Dennis McFadden. Masculinization of the mammalian cochlea. *Hearing research*, 252(1-2):37–48, 2009.
- Dennis McFadden and Frederic L Wightman. Audition: some relations between normal and pathological hearing. *Annual Review of Psychology*, 34(1):95–128, 1983.
- Dennis McFadden, John C Loehlin, and Edward G Pasanen. Additional findings on heritability and prenatal masculinization of cochlear mechanisms: click-evoked otoacoustic emissions. *Hearing Research*, 97(1-2):102–119, 1996.
- Dennis McFadden, Edward G Pasanen, Jessica Raper, Henry S Lange, and Kim Wallen. Sex differences in otoacoustic emissions measured in rhesus monkeys (*macaca mulatta*). *Hormones and behavior*, 50(2):274–284, 2006.
- Dennis McFadden, Glen K Martin, Barden B Stagner, and Mindy M Maloney. Sex differences in distortion-product and transient-evoked otoacoustic emissions compared. *The Journal of the Acoustical Society of America*, 125(1):239–246, 2009.
- David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. e1071: misc functions of the department of statistics, probability theory group (formerly: E1071), tu wien. r package version 1.6-8, 2017.
- Charles A Micchelli, Yuesheng Xu, and Haizhang Zhang. Universal kernels. *Journal of Machine Learning Research*, 7(Dec):2651–2667, 2006.
- Krikamol Muandet, Kenji Fukumizu, Francesco Dinuzzo, and Bernhard Schölkopf. Learning from distributions via support measure machines. In *Advances in neural information processing systems*, pages 10–18, 2012.
- Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, Bernhard Schölkopf, et al. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1-2):1–141, 2017.
- John Neter, Michael H Kutner, Christopher J Nachtsheim, and William Wasserman. *Applied linear statistical models*, volume 4. Irwin Chicago, 1996.
- Toshiaki O-Uchi, Jin Kanzaki, Yasuo Satoh, Shigemitsu Yoshihara, Akira Ogata, Yasuhiro Inoue, and Hiroyasu Mashino. Age-related changes in evoked otoacoustic emission in normal-hearing ears. *Acta Oto-Laryngologica*, 114(sup514):89–94, 1994.
- Junier Oliva, Willie Neiswanger, Barnabás Póczos, Jeff Schneider, and Eric Xing. Fast distribution to real regression. In *Artificial Intelligence and Statistics*, pages 706–714, 2014.

- G Paludetti, F Ottaviani, AR Fetoni, AA Zuppa, and G Tortorolo. Transient evoked otoacoustic emissions (teoaes) in new-borns: normative data. *International journal of pediatric otorhinolaryngology*, 47(3):235–241, 1999.
- Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *ICML 2016*, 2016.
- JO Pickles. Recent advances in cochlear physiology. *Progress in neurobiology*, 24(1):1–42, 1985.
- Barnabás Póczos, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. Distribution-free distribution regression. 2013.
- Cristian Preda. Regression models for functional data by reproducing kernel hilbert spaces methods. *Journal of statistical planning and inference*, 137(3):829–840, 2007.
- Beth A. Prieve, Tracy S. Fitzgerald, and Laura E. Schulte. Basic characteristics of click-evoked otoacoustic emissions in infants and children. *The Journal of the Acoustical Society of America*, 102(5):2860–2870, nov 1997a. doi: 10.1121/1.420341. URL <https://doi.org/10.1121%2F1.420341>.
- Beth A Prieve, Tracy S Fitzgerald, and Laura E Schulte. Basic characteristics of click-evoked otoacoustic emissions in infants and children. *The Journal of the Acoustical Society of America*, 102(5):2860–2870, 1997b.
- Rudolf Probst, Brenda L Lonsbury-Martin, and Glen K Martin. A review of otoacoustic emissions. *The Journal of the Acoustical Society of America*, 89(5):2027–2067, 1991.
- James Ramsay. Functional data analysis. *Encyclopedia of Statistics in Behavioral Science*, 2005.
- James Ramsay, Giles Hooker, and Spencer Graves. *Functional data analysis with R and MATLAB*. Springer Science & Business Media, 2009.
- James O Ramsay and Bernard W Silverman. *Applied functional data analysis: methods and case studies*. Springer, 2007.
- John A Rice. *Mathematical statistics and data analysis*. Cengage Learning, 2006.
- Joseph P Romano and Michael Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005.
- WLC Rutten. Evoked acoustic emissions from within normal and abnormal human ears: comparison with audiometric and electrocochleographic findings. *Hearing Research*, 2(3-4):263–271, 1980.
- YS Sininger and B Cone-Wesson. Asymmetric cochlear processing mimics hemispheric specialization. *Science*, 305(5690):1581–1581, 2004.
- Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A hilbert space embedding for distributions. In *International Conference on Algorithmic Learning Theory*, pages 13–31. Springer, 2007.
- Alex J Smola and Bernhard Schölkopf. *Learning with kernels*, volume 4. Citeseer, 1998.
- Vladimir G Spokoiny et al. Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6):2477–2498, 1996.

- Bharath Sriperumbudur, Kenji Fukumizu, and Gert Lanckriet. On the relation between universality, characteristic kernels and rkhs embedding of measures. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 773–780, 2010a.
- Bharath K Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Gert RG Lanckriet, and Bernhard Schölkopf. Injective hilbert space embeddings of probability measures. In *COLT*, volume 21, pages 111–122, 2008.
- Bharath K Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert RG Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(Apr):1517–1561, 2010b.
- Bharath K Sriperumbudur, Kenji Fukumizu, and Gert RG Lanckriet. Universality, characteristic kernels and rkhs embedding of measures. *Journal of Machine Learning Research*, 12(Jul):2389–2410, 2011.
- Ingo Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2(Nov):67–93, 2001.
- Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008.
- Zoltán Szabó, Bharath K Sriperumbudur, Barnabás Póczos, and Arthur Gretton. Learning theory for distribution regression. *The Journal of Machine Learning Research*, 17(1):5272–5311, 2016.
- Don C Teas. Auditory physiology: present trends. *Annual review of psychology*, 40(1):405–429, 1989.
- Terry Therneau, Beth Atkinson, Brian Ripley, and Maintainer Brian Ripley. Package `rpart`. Available online: [cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf](http://cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf) (accessed on 20 April 2016), 2018.
- Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 1st edition, 2008. ISBN 0387790519, 9780387790510.
- Andrea Vedaldi, Varun Gulshan, Manik Varma, and Andrew Zisserman. Multiple kernels for object detection. In *2009 IEEE 12th international conference on computer vision*, pages 606–613. IEEE, 2009.
- Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- Ward Whitt. Bivariate distributions with given marginals. *The Annals of statistics*, pages 1280–1289, 1976.
- HP Wit and HF Kahmann. Frequency analysis of stimulated cochlear acoustic emissions in monkey ears. *Hearing research*, 8(1):1–11, 1982.
- HP Wit and RJ Ritsma. Evoked acoustical responses from the human ear: some experimental results. *Hearing research*, 2(3-4):253–261, 1980.
- HP Wit, JC Langevoort, and RJ Ritsma. Frequency spectra of cochlear acoustic emissions (kemp-echoes). *The Journal of the Acoustical Society of America*, 70(2):437–445, 1981.
- E Zwicker. Delayed evoked oto-acoustic emissions and their suppression by gaussian-shaped pressure impulses. *Hearing research*, 11(3):359–371, 1983.

Eberhard Zwicker and Eberhard Schloth. Interrelation of different oto-acoustic emissions. *The Journal of the Acoustical Society of America*, 75(4):1148–1154, 1984.