

---

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université Abderrahmane Mira de Béjaïa



Faculté des Sciences Exactes  
Département de Mathématiques

# Mémoire de Fin de Cycle

EN VUE DE L'OBTENTION DU DIPLOME DE MASTER EN MATHÉMATIQUE

Option : Probabilités , Statistique et Application

## Thème :

Estimation par projection de la fonction de densité de probabilité.

Présenté par :

BENYAHIA MOHAND LAMINE

Soutenu le 02/juillet/2024 devant le jury composé de :

Présidente : Mme IDJIS K M.C.B - Université de Béjaïa  
Examineur : Mr BOURAINE M M.A.A - Université de Béjaïa  
Encadrant : Mme TABTI H M.C.B - Université de Béjaïa

Année Universitaire 2023/2024

# Remerciements

Je tiens à exprimer ici toute ma gratitude à Madame Hadjila TABTI pour l'honneur qu'elle m'a fait en assurant la direction du présent mémoire. Je la remercie pour ses précieux conseils et orientations.

Je suis heureux de pouvoir remercier Madame IDJIS pour avoir accepté de présider le jury de ma soutenance ainsi que Monsieur Mohand BOURAINE d'avoir accepté d'examiner ce travail.

Je n'oublierai pas de remercier ma famille qui m'a toujours encouragé(e) et soutenu(e) ainsi que toute personne ayant participé de près ou de loin à ce travail.

# Table des matières

<b>Table des matières</b>	<b>ii</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 Généralités sur les Variables Aléatoires Fonctionnelles</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Définition d'une variable aléatoire fonctionnelle . . . . .	3
1.3 Choix des données fonctionnelles et nécessité d'outils statistiques spécifiques . . . . .	4
1.4 Généralités et Rappels . . . . .	5
1.4.1 Quelques éléments d'analyse fonctionnelle . . . . .	5
1.4.2 Espaces de Hilbert . . . . .	7
1.4.3 Rappel sur les Processus Stochastiques . . . . .	7
1.4.4 Décomposition de Karhunen-Loève . . . . .	9
1.5 Représentation des fonctions dans une base . . . . .	10
1.5.1 Principe . . . . .	10
1.5.2 Bases classiques : . . . . .	10
<b>2 Données fonctionnelles et méthodes de lissage</b>	<b>14</b>
2.1 Introduction . . . . .	14
2.2 Lissage par Moindre Carrés . . . . .	15
2.2.1 Moindre carrés ordinaire : . . . . .	15
2.2.2 Choix de la dimension de l'espace d'approximation . . . . .	17
2.3 Lissage par moindres carrés pénalisés : . . . . .	17
2.3.1 Principe . . . . .	17
2.3.2 Choix du parametre de lissage . . . . .	18
<b>3 Estimation non paramétrique de la densité de probabilité par la méthode des fonctions orthogonales</b>	<b>20</b>
3.1 Description de la méthode . . . . .	20
3.1.1 Propriétés de l'estimateur . . . . .	21
3.2 Estimateurs associés aux fonctions trigonométriques . . . . .	22
3.2.1 Estimateur de Dirichlet . . . . .	22

---

3.2.2	Propriété . . . . .	22
3.3	Estimateur associé a la base des Ondelette . . . . .	24
3.3.1	Estimateur de densité de Haar . . . . .	24
3.3.2	Propriétés . . . . .	24
<b>4</b>	<b>APPLICATION</b>	<b>26</b>
4.1	Introduction . . . . .	26
4.2	Logiciel utilisé : . . . . .	26
4.2.1	Caractéristiques Principales . . . . .	26
4.2.2	Utilisation de R pour l'Estimation Non Paramétrique . . . . .	27
4.3	Données Utilisées . . . . .	27
4.4	Algorithme . . . . .	27
4.4.1	Calcule des moyennes et représantation . . . . .	27
4.4.2	Hypothese sur la nature des données . . . . .	28
4.4.3	Calcule de la densité vrai et représantation . . . . .	28
4.4.4	Estimation de la densité . . . . .	29
4.4.5	Calcule du MISE : . . . . .	29
4.5	Conclusion . . . . .	30
	<b>Conclusion générales et perspectives</b>	<b>31</b>
	<b>Bibliographie</b>	<b>33</b>

## Résumé

L'estimation non paramétrique de la densité de probabilité est un domaine clé en statistique moderne. Dans ce mémoire, nous explorons des méthodes d'estimation utilisant des fonctions orthogonales. Nous commençons par une présentation générale des variables aléatoires fonctionnelles et des bases théoriques nécessaires, incluant les espaces de Hilbert et les processus stochastiques. Ensuite, nous décrivons les méthodes d'estimation non paramétrique de la densité en utilisant des fonctions orthogonales comme les fonctions trigonométriques et les ondelettes de Haar. Nous détaillons les propriétés de ces estimateurs, notamment le biais et la variance. Enfin, nous illustrons ces méthodes par une application pratique sur des données météorologiques, démontrant l'efficacité des approches proposées.

**Mots clés :** Estimation non paramétrique, densité de probabilité, fonctions orthogonales, données fonctionnelles, analyse statistique, données météorologiques.

## **Abstract**

Non-parametric density estimation is a key area in modern statistics. In this thesis, we explore estimation methods using orthogonal functions. We begin with a general introduction to functional random variables and the necessary theoretical bases, including Hilbert spaces and stochastic processes. Then, we describe non-parametric density estimation methods using orthogonal functions such as trigonometric functions and Haar wavelets. We detail the properties of these estimators, including bias and variance. Finally, we illustrate these methods with a practical application on meteorological data, demonstrating the effectiveness of the proposed approaches.

**Key words :** Non-parametric estimation, probability density, orthogonal functions, functional data, statistical analysis, meteorological data.

# Introduction Generale

Estimation de la densité de probabilité est un domaine crucial en statistique, essentiel pour comprendre la distribution sous-jacente des données. Historiquement, les méthodes paramétriques ont été les premières approches développées pour cette tâche. Ces méthodes reposent sur l'hypothèse que les données suivent une distribution spécifique, définie par un petit nombre de paramètres. Par exemple, l'estimation des paramètres de la distribution normale par Karl Pearson à la fin du 19ème siècle est une des premières applications des méthodes paramétriques. Cette approche présente l'avantage de simplifier le problème d'estimation, mais elle impose des contraintes fortes sur la forme de la distribution.

Cependant, les méthodes paramétriques ont des limitations significatives lorsqu'il s'agit de modéliser des distributions complexes ou inconnues. Les distributions réelles des données peuvent souvent s'écarter de manière substantielle des distributions paramétriques supposées, rendant ces méthodes inappropriées pour certaines applications. Pour surmonter ces limitations, les méthodes non paramétriques ont été développées, offrant une flexibilité accrue en ne supposant aucune forme prédéfinie pour la densité.

L'approche non paramétrique de l'estimation de la densité a gagné en popularité grâce aux travaux de chercheurs comme Emanuel Parzen et Murray Rosenblatt dans les années 1950. Ces méthodes estiment directement la densité de probabilité à partir des données, sans imposer de contraintes sur la forme de la distribution. Parmi les méthodes non paramétriques les plus connues, on trouve l'estimation par histogramme et l'estimation par noyau. L'estimation par histogramme, bien que simple et intuitive, souffre de plusieurs inconvénients tels que la sensibilité à la largeur des intervalles, ce qui peut affecter la précision et la régularité de la densité estimée. En revanche, l'estimation par noyau, introduite par Parzen, offre une estimation plus lisse et plus précise, en utilisant une fonction noyau pour lisser les données.

Dans ce mémoire, nous explorerons une autre alternative prometteuse : la méthode des fonctions orthogonales pour l'estimation non paramétrique de la densité de probabilité. Cette méthode repose sur l'utilisation de bases orthogonales, comme les polynômes de Legendre ou les séries de Fourier, pour représenter la densité de manière flexible et précise. En utilisant cette approche, nous pouvons capturer les caractéristiques essentielles des données sans imposer de contraintes rigides sur leur distribution. Nous démontrerons l'efficacité de cette méthode à travers des applications pratiques sur des données réelles, notamment des données météorologiques.

Ce mémoire se concentre sur l'estimation non paramétrique de la densité de probabilité à travers l'utilisation des méthodes de projection. L'objectif principal est de développer une approche basée sur les fonctions orthogonales pour estimer la densité de probabilité de manière précise et efficace. En particulier, nous nous intéressons aux techniques

de lissage et de sélection de modèles qui permettent d'améliorer la qualité des estimations obtenues.

Nous proposons également une application pratique de ces méthodes sur des données réelles, en utilisant des données météorologiques pour illustrer l'efficacité et la pertinence de notre approche. Cette application permet non seulement de démontrer la robustesse de notre méthode, mais aussi d'explorer les dynamiques climatiques à travers l'analyse statistique des données de température.

Ce mémoire est structuré en plusieurs chapitres, chacun abordant un aspect spécifique de l'estimation par projection de la densité de probabilité et de son application.

Le premier chapitre offre une revue exhaustive des généralités et rappels sur les variables aléatoires fonctionnelles, éléments indispensables à la compréhension de l'estimation non paramétrique de la densité. Nous y discutons en détail des propriétés et des caractéristiques des variables aléatoires fonctionnelles, en mettant en lumière leur rôle crucial dans le processus d'estimation de la densité de probabilité.

Dans le second chapitre, nous présenterons les fondements théoriques de l'estimation par projection, en détaillant les propriétés des bases orthogonales et leur utilisation pour la représentation des densités de probabilité. Ce chapitre inclut également une discussion sur les techniques de lissage et les critères de sélection de modèles.

Le troisième chapitre explore l'utilisation des modèles graphiques pour capturer les dépendances entre les variables aléatoires fonctionnelles. Nous y détaillons la construction de ces modèles et leur intégration dans le cadre de l'estimation par projection. De plus, nous présentons les estimateurs basés sur les fonctions orthogonales déjà connues, en fonction de la base orthogonale choisie, ainsi que leurs propriétés.

Dans le quatrième chapitre, nous appliquons les méthodes développées aux données météorologiques de plusieurs villes d'Algérie. Nous utilisons les fonctions orthogonales pour estimer la densité des températures sur une période d'un an et comparons les résultats obtenus à la densité vraie des données. Cette application illustre la pertinence et l'efficacité de notre approche pour l'analyse des données réelles.



# Généralités sur les Variables Aléatoires Fonctionnelles

## 1.1 Introduction

Le modèle communément utilisé en analyse de données fonctionnelles, et convenant en particulier pour tous les exemples précédemment cités est le suivant : on considère que les courbes sont des réalisations de processus stochastiques comportant des trajectoires relativement régulières (lisses). Dans la suite de ce chapitre, on note  $(\Omega, \mathcal{A}, P)$  un espace probabilisé ( $\Omega$  est un ensemble,  $\mathcal{A}$  une tribu sur cet ensemble, et  $P$  une mesure de probabilité sur  $\mathcal{A}$ ). On note également  $F$  un espace de fonctions (dont on supposera que c'est au moins un espace de Banach séparable, c'est-à-dire un espace vectoriel normé complet, contenant un sous ensemble dense dénombrable).

## 1.2 Définition d'une variable aléatoire fonctionnelle

Une variable aléatoire est dite variable aléatoire fonctionnelle si elle prend ses valeurs dans un espace vectoriel de dimension infinie. Typiquement, il s'agit donc d'une application mesurable  $X : \Omega \rightarrow F$ . Une donnée fonctionnelle est alors une réalisation de la variable  $X$ .

Si l'espace  $F$  est constitué de fonctions définies sur un ensemble  $T$  (par exemple l'intervalle  $[0, 1]$ ) et à valeurs réelles, on peut voir une variable fonctionnelle  $X$  comme une application  $X : \Omega \times T \rightarrow \mathbb{R}$ , telle que

- pour tout  $\omega \in \Omega$  fixé, la fonction  $X(\omega, \cdot) : T \rightarrow \mathbb{R}$  est une trajectoire de  $X$ ,
- pour tout  $t \in T$  fixé, la variable  $X(\cdot, t) : \Omega \rightarrow \mathbb{R}$  est une variable aléatoire réelle.

Si l'ensemble  $T$  est inclus dans  $\mathbb{R}$ , la variable  $X$  est une courbe aléatoire ; lorsque  $T$  est de dimension supérieure, on a un processus aléatoire plus complexe, par exemple une surface aléatoire lorsque  $T$  est inclus dans  $\mathbb{R}^2$  (une image typiquement). Comme habituellement en probabilités et en statistique, on omettra les arguments  $\omega$  et  $t$  dans la suite, s'il n'y a pas de confusion possible, et on confondra les variables et leurs réalisations.

Dans la suite de ce chapitre, on considérera souvent un échantillon  $X_1, \dots, X_n$  ( $n \in \mathbb{N} \setminus \{0\}$  est la taille de l'échantillon) de données fonctionnelles indépendantes et identiquement distribuées selon la loi de  $X$ . Notons qu'il est également possible dans ce cadre fonctionnel de considérer des données dépendantes (séries temporelles fonctionnelles par exemple), mais nous ne traiterons pas ce cas.

### 1.3 Choix des données fonctionnelles et nécessité d'outils statistiques spécifiques

On n'observe en réalité jamais une courbe aléatoire ou un processus  $X$  dans son ensemble : cela supposerait en effet d'une part que l'on dispose d'instruments de mesure avec une vitesse d'enregistrement infinie, et d'autre part que l'on est capable de stocker un nombre infini (non dénombrable) de valeurs. Une donnée fonctionnelle se présente donc en fait tout d'abord sous forme vectorielle : elle est constituée d'un certain nombre de valeurs discrètes qui ont été mesurées sur une grille suffisamment fine, et enregistrées. Considérons des fonctions d'un ensemble  $T$  dans  $\mathbb{R}$ . Ainsi, on n'observe pas intégralement l'ensemble  $\{X(t), t \in T\}$ , mais un ensemble  $\{X(t_1), \dots, X(t_p)\}$ , où  $\{t_1, \dots, t_p\} \subset T$  est une grille de discrétisation (une suite ordonnée de valeurs). Dans le cas d'un échantillon de données, on a généralement accès à  $n$  vecteurs  $(X_i(t_{i,1}), \dots, X_i(t_{i,p}))$ , où  $\{t_{i,1}, \dots, t_{i,p}\}$  représente toujours la grille, pour  $i \in \{1, \dots, n\}$  : cette grille peut, ou non, être identique pour toutes les courbes  $X_i$  de l'échantillon.

Ce type de données n'est pas nouveau, et a longtemps été étudié avec les techniques classiques de la statistique multivariée : les données sont alors vues comme des vecteurs aléatoires de  $\mathbb{R}^p$ . L'apport de la statistique pour données fonctionnelles est de traiter ces données comme des réalisations de fonctions aléatoires (en théorie, on pourrait obtenir des points aussi rapprochés que l'on veut, c'est-à-dire des mesures sur une grille extrêmement fine). Les avantages sont les suivants :

1. **Prise en compte de grilles d'enregistrement possiblement différentes** : Il arrive que les données récoltées ne le soient pas sur la même grille, en pratique. C'est par exemple le cas lorsqu'on a des données incomplètes, avec des observations manquantes. L'analyse de données fonctionnelles propose un cadre théorique rigoureux permettant de prendre en compte ce problème, et permet d'atténuer les effets de la non-correspondance des instants d'observation : si on dispose de données observées par exemple à des instants différents, une approximation continue permet d'évaluer la variable aux mêmes points pour tous les individus.
2. **Prise en compte des problématiques liées à la grande dimension** : Il arrive de plus en plus souvent, du fait de l'amélioration des appareils de mesures en matière de recueil et stockage des données, que la fréquence de discrétisation des courbes, c'est-à-dire la taille  $p$  de la grille, soit très grande, et surtout très grande devant la taille  $n$  de l'échantillon elle-même : on note généralement  $p \geq n$ . Les méthodes classiques de statistique multivariée se sont alors montrées inadéquates dans ce contexte, où paradoxalement, l'afflux de données aboutit à une détérioration des résultats classiques. C'est la fameuse "malédiction de la dimension" (curse of dimensionality). Pour y faire face, des outils propres aux problèmes de la grande dimension se sont développés ces dernières années, parmi lesquelles les méthodes et modèles de la statistique pour données fonctionnelles.

3. **Prise en compte de la structure intrinsèque des données :** (et en particulier de la régularité du phénomène observé). Considérer qu'on peut traiter les données explicitées sous forme discrètes comme des fonctions permet de prendre en compte la nature continue intrinsèque du phénomène observé, celle de fonctions aléatoires, et permet de faire intervenir également leurs régularités. Une corrélation importante existe généralement entre deux observations rapprochées d'un même phénomène continu : si  $X$  mesure un phénomène continu,  $X(t_j)$  et  $X(t_{j+1})$  sont nécessairement liés l'un à l'autre et peu susceptibles d'être très différents. Ceci incite à considérer par exemple des opérations de dérivation, pour étudier la régularité du phénomène, opérations qui n'auraient pas de sens dans un cadre vectoriel classique. Grâce à l'utilisation de données fonctionnelles, on pourra par exemple, pour un phénomène donné, étudier non seulement le phénomène lui-même mais tirer également parti de sa vitesse ou son accélération, en considérant les dérivées premières et secondes de  $X$ . L'étude des dérivées permet également d'analyser les aspects géométriques des variables. On peut aussi introduire d'autres informations a priori sur les courbes étudiées. Lors de la modélisation, on pourra par exemple prendre en compte des phénomènes de périodicité.

## 1.4 Généralités et Rappels

Il n'est pas question, pour une introduction aux données fonctionnelles, de faire un cours complet de topologie, puis d'analyse fonctionnelle. On se contentera donc de quelques éléments très restreints. Pour plus de détails, on pourra se référer à Briane et Pages (2000) pour tout ce qui concerne l'intégration, la théorie de la mesure, les espaces  $L^p$ , et Hirsch et Lacombe (1997) ou Brezis (1983) pour l'analyse fonctionnelle à proprement parler et en particulier la théorie des opérateurs.

### 1.4.1 Quelques éléments d'analyse fonctionnelle

#### Espaces vectoriels normés (Rudin 1991)

Nous avons défini une variable aléatoire fonctionnelle  $X$  comme une variable à valeurs dans un espace de fonctions  $F$ . Il va être fondamental de considérer uniquement des espaces vectoriels normés, pour pouvoir déterminer des développements de la fonction aléatoire  $X$  dans une base donnée. Rappelons qu'un espace vectoriel (réel)  $E$  est dit normé s'il est muni d'une application  $\|\cdot\| : E \rightarrow \mathbb{R}_+$  qui est une norme, c'est-à-dire qui vérifie les 3 hypothèses suivantes :

1. **Séparation** :  $\forall x \in E, \|x\| = 0 \Rightarrow x = 0$ ,
2. **Homogénéité** :  $\forall x \in E, \lambda \in \mathbb{R}, \|\lambda x\| = |\lambda| \|x\|$ ,
3. **Inégalité triangulaire** :  $\forall x, y \in E, \|x + y\| \leq \|x\| + \|y\|$ .

On peut définir plusieurs normes sur un même espace vectoriel. Si l'espace  $E$  est de dimension finie, toutes les normes sont équivalentes : si  $\|\cdot\|_a$  et  $\|\cdot\|_b$  sont deux normes sur  $E$ , il existe deux constantes  $c, C > 0$  telles que pour tout  $x \in E, c\|x\|_a \leq \|x\|_b \leq C\|x\|_a$ . Ce n'est plus le cas en dimension infinie, en particulier pour les espaces de fonctions. On rappelle également la notion de convergence d'une suite dans un espace vectoriel normé  $(E, \|\cdot\|)$  : une suite  $(x_n)_{n \in \mathbb{N}}$  de points de  $E$  a pour limite  $x \in E$  dans  $(E, \|\cdot\|)$ , si  $\lim_{n \rightarrow \infty} \|x_n - x\| = 0$ . En dimension finie, si une suite de  $E$  converge pour une norme donnée, elle converge au sens de toutes les autres. Ce n'est pas le cas en dimension infinie, où il est donc essentiel de spécifier la norme utilisée.

Toute suite convergente dans  $(E, \|\cdot\|)$  est de Cauchy, mais la réciproque n'est pas toujours vraie. Un espace vectoriel normé  $(E, \|\cdot\|)$  est dit complet si toute suite de Cauchy dans  $(E, \|\cdot\|)$  converge. On parle alors d'espace de Banach. Tout espace de dimension finie est complet. En dimension infinie, il existe des espaces non complets.

Enfin, un espace vectoriel normé est dit séparable s'il contient un sous-ensemble dénombrable dense, c'est-à-dire s'il n'est pas "trop grand" (ce qui ne veut pas dire qu'il n'est pas de dimension infinie!).

### Espace normé de dimension finie (Rudin 1991)

L'exemple le plus classique d'espace vectoriel normé de dimension finie est  $\mathbb{R}^d$ , muni de la norme dite euclidienne usuelle :

$$\forall x = (x_1, \dots, x_d) \in \mathbb{R}^d, \|x\| = \left( \sum_{j=1}^d x_j^2 \right)^{1/2}.$$

### Espaces fonctionnels de dimension infinie (Brezis 2011)

Pour simplifier, on se contente de considérer des fonctions définies sur  $[0, 1]$  et à valeurs réelles, même s'il est plus fréquent de s'intéresser aux fonctions définies sur  $\mathbb{R}^d$ .

### Espace des fonctions continues sur un compact (Rudin 1991)

On note  $C_0([0, 1])$  l'espace des fonctions continues sur  $[0, 1]$ . C'est un espace de Banach séparable pour la norme infinie, définie par :  $\forall x \in C_0([0, 1]), \|x\|_\infty = \sup_{t \in [0, 1]} |x(t)|$ .

### Espace des fonctions continument dérivables (Brezis 1983)

On note plus généralement, pour tout entier  $m \geq 1$ ,  $C^m([0, 1]) = \{x : [0, 1] \rightarrow \mathbb{R}, x^{(m)} \in C_0([0, 1])\}$ . Il est également possible de munir cet espace d'une norme qui en fait un Banach. Dans le cas où  $m = 1$  par exemple, on pose  $\|x\| = \|x\|_\infty + \|x'\|_\infty$ .

### Espaces $L^p$ (Evans 1992)

Pour simplifier, on définira l'espace  $L^p([0, 1])$  comme l'espace des fonctions de puissance  $p$ -ième intégrable sur  $[0, 1]$ , pour  $1 \leq p < \infty$  :

$$L^p([0, 1]) = \left\{ x : [0, 1] \rightarrow \mathbb{R}, \int_{[0, 1]} |x(t)|^p dt < \infty \right\}.$$

Le théorème de Riesz-Fischer prouve que les espaces  $L^p$ , munis de la norme  $\|x\|_p = \left( \int_{[0, 1]} |x(t)|^p dt \right)^{1/p}$ ,  $x \in L^p([0, 1])$ , sont des espaces de Banach séparables.

### 1.4.2 Espaces de Hilbert

Nous allons très vite nous restreindre à des espaces de Banach particulier dont nous rappelons la définition et quelques propriétés ici. On appelle d'abord espace préhilbertien réel un espace vectoriel  $E$  sur  $\mathbb{R}$  muni d'un produit scalaire, c'est à dire d'une forme  $\langle \cdot, \cdot \rangle$ , application de  $E \times E$  dans  $\mathbb{R}$  qui est

1. bilinéaire : linéaire par rapport à chacune de ses variables,
2. symétrique :  $\forall x, y \in E, \langle x, y \rangle = \langle y, x \rangle$ ,
3. positive :  $\forall x \in E, \langle x, x \rangle \geq 0$ ,
4. définie :  $\forall x \in E, \langle x, x \rangle = 0 \Rightarrow x = 0$ .

Deux éléments  $x, y$  de  $(E, \langle \cdot, \cdot \rangle)$  sont dits orthogonaux si  $\langle x, y \rangle = 0$ . Tout espace préhilbertien  $(E, \langle \cdot, \cdot \rangle)$  est un espace vectoriel normé : en posant, quelque soit  $x \in E, \|x\| = \sqrt{\langle x, x \rangle}$ , on définit bien une norme. Deux éléments de  $E$  sont dits orthonormés s'ils sont orthogonaux et de norme 1.

Un espace de Hilbert est un espace préhilbertien qui est complet. C'est donc un espace de Banach dans lequel la norme découle d'un produit scalaire par la formule ci-dessus (Conway 1990).

#### Exemples utiles

1. **Espace euclidien  $\mathbb{R}^d$**  : L'espace vectoriel  $\mathbb{R}^d$  est un espace de Hilbert de dimension finie, c'est-à-dire un espace euclidien, pour le produit scalaire

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^d x_j y_j, \quad (\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d, \mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d)$$

. La norme euclidienne est une norme découlant du produit scalaire.

2. **Espace  $L^2$**  : L'espace  $L^2([0, 1])$  des fonctions de carré intégrable sur  $[0, 1]$ , muni de  $\langle x, y \rangle = \int_0^1 |x(t)y(t)| dt$ , pour  $x, y \in L^2([0, 1])$ , est un espace de Hilbert.

ç Le résultat le plus important qui nous sera utile est le suivant :

**Théorème 1** Soit  $(H, h, \cdot, i)$  un espace de Hilbert séparable.

1.  $H$  admet une (des) base(s) hilbertienne(s) dénombrable(s), c'est-à-dire une famille  $\{\varphi_j, j \in \mathbb{N}\}$  d'éléments orthonormés de  $H$  totale, au sens où l'adhérence de l'espace vectoriel engendré par cette famille est égale à l'espace  $H$  tout entier. Cela signifie que tout élément de  $H$  se décompose de façon unique sous la forme :

$$\forall x \in H, x = \sum_{j \in \mathbb{N}} \langle x, \varphi_j \rangle \varphi_j.$$

2. Toute forme linéaire sur  $H, f : H \rightarrow \mathbb{R}$ , peut se mettre sous la forme  $f = \langle \cdot, x \rangle$  pour un élément  $x$  de  $H$ . On dit que  $H$  est isomorphe à son dual  $H^*$  (Fréchet-Riesz 1907).

### 1.4.3 Rappel sur les Processus Stochastiques

La question posée dans cette section est celle de l'extension des notions d'espérance et de matrice de covariance d'un vecteur aléatoire aux variables à valeurs dans un espace de dimension infinie. On suppose dans la suite que  $F$  est un espace de Banach, plus précisément un espace de fonctions définies sur un ensemble  $T$  et à valeurs réelles, dont on note

$\|\cdot\|$  la norme, et muni de la tribu borélienne, c'est-à-dire la tribu associée à la norme,  $(\Omega, \mathcal{A}, P)$  un espace probabilisé, et  $X$  une variable aléatoire fonctionnelle sur  $\Omega$ , à valeurs dans  $F$ .

### Espérance fonctionnelle

**Définition 1** Lorsque cela a un sens, on peut définir l'espérance de  $X$  comme la fonction non aléatoire  $(E[X])(\cdot)$  (élément de  $F$ ) définie point par point par

$$(E[X])(t) = E[X(t)] = \int_{\Omega} X(t, \omega) dP(\omega),$$

où  $\Omega$  est l'espace des événements,  $P$  est la mesure de probabilité, et  $X$  est une variable aléatoire fonctionnelle définie sur  $\Omega \times T$  (Ramsey-Silverman 2005).

**Propriétés classiques de l'espérance :** Par exemple, pour toute variable aléatoire  $X$  définie sur un espace de Hilbert  $F$ , on a les propriétés suivantes :

1.  $\|E[X]\| \leq E[\|X\|]$
2.  $\|E[X]\|^2 \leq E[\|X\|^2]$

Lorsque  $F$  est un espace de Hilbert muni du produit scalaire  $\langle \cdot, \cdot \rangle$ , on a plus généralement

3.  $E[\langle X, f \rangle] = \langle E[X], f \rangle$   
pour tout  $f \in F$ .

### Opérateur et Fonction Covariance

Dans le cadre multivarié, la covariance a une forme matricielle : l'extension naturelle à la dimension infinie est donc un opérateur. On se restreint à présent au cas où  $F$  est un espace de Hilbert.

**Définition 2** Supposons que  $E[\|X\|^2] < \infty$ . L'opérateur de covariance de  $X$  est défini par

$$\Gamma : f \in F \mapsto \Gamma f = E[(X - E[X]), f(X - E[X])].$$

Il est à valeurs dans  $F$  (Ramsey-Silverman 2005).

On a, dans le cas hilbertien, le résultat suivant, que l'on pourra omettre si l'on n'est pas familier de la théorie des opérateurs, pour n'en retenir que les conséquences

**Proposition 1** Supposons que  $X$  est une variable aléatoire à valeurs dans un espace de Hilbert séparable  $F$  telle que  $E[\|X\|^2] < \infty$ . Alors, son opérateur de covariance  $\Gamma$  est

1. autoadjoint :  $\forall f, g \in F, \langle \Gamma f, g \rangle = \langle f, \Gamma g \rangle$ ,
2. de Hilbert-Schmidt : il existe une base hilbertienne  $(\phi_j)_{j \in \mathbb{N}}$  telle que  $\sum_{j \in \mathbb{N}} \|\Gamma \phi_j\|^2 < \infty$  (Ramsey-Silverman 2005).

on en déduit ,

**Corollaire 1** Avec les hypothèses et notations de la proposition précédente, l'opérateur de covariance  $\Gamma$  de  $X$  est ainsi

1. compact (l'image de la boule unité de  $F$  par  $\Gamma$  est relativement compacte),

2. diagonalisable en base orthonormée.

(Brezis 2011)

Une autre quantité que l'opérateur de covariance est la fonction covariance, fréquemment utilisée en théorie des processus stochastiques. Pour simplifier (encore), on suppose que  $F$  est l'espace  $L^2([0, 1])$ .

**Définition 3** Soit  $X$  une variable aléatoire à valeurs dans  $L^2([0, 1])$ . La fonction de covariance de  $X$  est l'application

$$C : [0, 1]^2 \rightarrow \mathbb{R}, \quad (s, t) \mapsto C(s, t) = E[X(t)X(s)].$$

(Ramsey-Silverman 2005)

**Proposition 2** Soit  $X$  une variable aléatoire à valeurs dans  $L^2([0, 1])$ , d'opérateur de covariance  $\Gamma$  et de fonction de covariance  $C$ . Alors, pour tout  $f \in L^2([0, 1])$  et tout  $t \in [0, 1]$ , on a

$$(\Gamma f)(t) = \int_0^1 C(s, t)f(s) ds.$$

Ainsi,  $\Gamma$  est un opérateur à noyau, de noyau la fonction de covariance (Ramsey-Silverman 2005).

#### 1.4.4 Décomposition de Karhunen-Loève

On considère ici  $X$  une variable aléatoire à valeurs dans un espace de Hilbert  $(F, h, \cdot, \cdot)$ , d'opérateur de covariance  $\Gamma$ . Nous avons vu que  $\Gamma$  est diagonalisable (Corollaire 1). Ainsi, il existe une base hilbertienne de  $F$ , formée de vecteurs propres  $(\psi_j)_{j \in \mathbb{N}}$  de  $\Gamma$  associés aux valeurs propres  $(\lambda_j)_{j \in \mathbb{N}}$  rangées par ordre décroissant (c'est-à-dire  $\Gamma\psi_j = \lambda_j\psi_j$ ). Ces valeurs propres sont positives. On peut alors développer  $X$  dans cette base hilbertienne (voir Théorème 1).

**Définition 4** On appelle développement de Karhunen-Loève d'une variable  $X$  à valeurs dans un espace de Hilbert telle que  $E[\|X\|^2] < \infty$ , son développement dans la base de fonctions propres de l'opérateur de covariance associé. Avec les notations précédentes, on obtient l'écriture en série aléatoire suivante :

$$X = E[X] + \sum_{j=1}^{\infty} \langle X, \psi_j \rangle \psi_j = E[X] + \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_j \psi_j,$$

où :  $\xi_j = \langle X, \psi_j \rangle / \sqrt{\lambda_j}$  (Ramsey-Silverman 2005).

## 1.5 Représentation des fonctions dans une base

### 1.5.1 Principe

On se place dans le cas de variables aléatoires à valeurs dans  $F = L^2(T)$ , où  $T$  est un intervalle de  $\mathbb{R}$ . Les sections précédentes montrent l'importance, pour une variable fonctionnelle, de la décomposition de Karhunen-Loève. Celle-ci étant en général inconnue, il va être important de pouvoir approcher les variables fonctionnelles comme combinaisons linéaires de fonctions d'une base donnée. Rappelons qu'un système de fonctions  $(\phi_j)_{j \in \mathbb{N}}$  forme une base hilbertienne s'il est composé de fonctions orthonormales (donc linéairement indépendantes), qui ont la propriété suivante : en prenant une combinaison linéaire d'un nombre suffisamment grand de ces fonctions, on peut approcher n'importe quelle autre fonction de l'espace  $F$ . Plus précisément, si  $X$  est une variable à valeurs dans  $F = L^2(T)$ , on a l'égalité suivante, dans  $L^2(T)$ ,

$$X = \sum_{j=1}^{\infty} \theta_j \phi_j,$$

la convergence de la série ayant lieu dans  $L^2(T)$ . L'idée naturelle pour approcher  $X$  est alors de tronquer la série à un certain niveau  $D$ , et d'estimer les coefficients : il s'agit donc d'approcher  $X$  par  $\tilde{X}$ , définie par

$$\tilde{X}(t) = \sum_{j=1}^D \tilde{\theta}_j \phi_j(t), \quad t \in T.$$

On obtient ainsi une approximation appartenant à un sous-espace de dimension finie  $D$  de  $L^2(T)$  (l'espace vectoriel engendré par les  $D$  premières fonctions de la base choisie).

En notant  $\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_D)^t$  le vecteur (colonne) des coefficients et  $\Phi(t) = (\phi_1(t), \dots, \phi_D(t))$  le vecteur dont les éléments sont les  $D$  premiers éléments de la base évalués en  $t$ , on a la représentation matricielle suivante :

$$\tilde{X}(t) = \tilde{\theta} \Phi(t)^t,$$

utile pour l'implémentation. Lorsque l'on souhaite approcher une fonction, les coefficients  $\theta$  sont bien évidemment à estimer, mais ce ne sont pas les seuls paramètres inconnus : il faut voir aussi  $D$ , la dimension de l'espace d'approximation, comme un paramètre à choisir (voir Chapitre 3). Le choix de la base  $(\phi_j)$  dépend des hypothèses a priori sur la variable à étudier. On en présente brièvement quelques une, facilement implementable sous logiciel R. En particulier, le choix de la base peut dépendre du fait qu'on veut ou non estimer les dérivées de  $X$ . Une représentation précise et fidèle de  $X$  à l'aide de nombreuses oscillations peut par exemple mener à une très mauvaise représentation des dérivées. (Green-Silverman 1994)

### 1.5.2 Bases classiques :

#### 1. Base de Fourier :

La base de Fourier, ou base trigonométrique, est probablement la plus connue des bases hilbertiennes, et la plus appropriée pour approcher des courbes (aléatoires) au comportement périodique (comme les phénomènes d'évolution de température par exemple). Elle est définie de la manière suivante :



quel que soit  $t$

$$\psi_0(t) = 1, \quad \psi_{2k-1}(t) = \sin(k\omega t), \quad \psi_{2k}(t) = \cos(k\omega t)$$

pour  $k \in \mathbb{N} \setminus \{0\}$

Le paramètre  $\omega$  détermine la période de la fonction, qui est donnée par  $2\pi/\omega$ .

Les coefficients associés au développement d'une fonction dans cette base sont appelés les **coefficients de Fourier**. Les dix premières fonctions de la base de Fourier, avec  $\omega = 2\pi$ , sont particulièrement adaptées pour la reconstruction de fonctions très régulières, présentant une courbure à peu près uniforme en tous points. Elle ne convient pas pour des fonctions reflétant des discontinuités, en raison, entre autres, d'effets de bord importants (oscillations d'amplitude importante). On parle de phénomène de Gibbs pour les effets de bords au voisinage des discontinuités de la fonction. On préférera alors utiliser des bases dites "localisées".

## 2. Base polynomiale et base de B-splines :

La base des monômes  $\phi_j(t) = (t - \omega)^j$ ,  $j \geq 0$ , où  $\omega$  est le paramètre de translation, souvent choisi comme le milieu de l'intervalle  $T$ , a longtemps été la référence pour la reconstruction de fonctions non-périodiques, en raison de la simplicité d'estimation des coefficients.

Cependant, la base des monômes a progressivement été remplacée, d'abord par des bases de polynômes par morceaux (définies par exemple à partir des polynômes de Legendre), que nous ne détaillerons pas (voir par exemple Comte 2015), ou des bases de splines. Il serait hors de propos de faire ici un cours entier sur les splines, nous donnons ici juste quelques éléments sur la notion de fonction spline, et une manière classique d'en construire. Pour une étude détaillée, on pourra se référer à de Boor (2001).

**Définition 5** Une spline sur un intervalle  $T = [a, b]$  est une fonction possédant les propriétés suivantes :

- elle se caractérise par la donnée de  $L$  sous-intervalles de  $T$  délimités par des points appelés noeuds ("knots") ou points de rupture ("breakpoints")  $\tau_0 = a \leq \tau_1 \leq \dots \leq \tau_L = b$ , non nécessairement répartis régulièrement dans  $T$  ;
- sur chaque sous-intervalle, la spline est un polynôme, d'ordre  $m/2$  ;
- les liaisons entre les différents polynômes aux points de rupture sont régulières, et, plus précisément, pour une spline définie avec des polynômes d'ordre  $m$ , les dérivées doivent être continues sur  $T$  jusqu'à l'ordre  $m - 2$ .

Pour un Exemple , on appelle spline cubique une spline d'ordre 4 (Boor 2001) .

Ainsi, plus les splines ont un ordre élevé, plus elles sont régulières. Et plus on augmente le nombre de noeuds, plus on gagne en flexibilité et précision : pour approcher une fonction avec des splines, on aura tendance à placer plus de noeuds aux endroits de  $T$  où la fonction semble avoir le comportement le plus complexe. On notera également que toute combinaison linéaire de fonctions spline est encore une fonction spline. Retenons, pour la suite, qu'une fonction spline est polynomiale par morceaux, avec conditions de continuité sur la fonction et ses dérivées aux jointures.

**remarque :** Les notions de “noeuds” et de “points de rupture” ne coïncident en fait pas tout à fait. Il est possible d’avoir plusieurs “points de rupture” qui coïncident. Le terme de “point de rupture” se réfère précisément aux valeurs des noeuds répétées une seule fois, tandis que le terme “noeuds” se réfère aux points de rupture, répétées avec multiplicité : un point de rupture peut être égal à plusieurs noeuds

**Définition 6** *On appellera base de spline d’ordre  $m$  et de séquence de noeuds  $\tau$  une famille de fonctions vérifiant les propriétés suivantes :*

- (a) *chaque fonction de base est elle-même une fonction spline (toute combinaison linéaire de ces fonctions est donc encore une fonction spline) ;*
- (b) *toute spline d’ordre  $m$  et de séquence de noeuds  $\tau$  peut s’exprimer comme combinaison linéaire de ces fonctions de base ;*
- (c) *les fonctions de bases sont linéairement indépendantes (pas nécessairement orthonormées).*

(Boor 2001 )

**B-splines :** Il existe plusieurs systèmes classiques de telles bases. La plus célèbre et fréquemment utilisée est sans doute celle introduite par de Boor (2001), appelée système de B-splines.

Un tel système est entièrement caractérisé par :

- un ordre  $m$  ou, de manière équivalente, le degré maximal  $m - 1$  des morceaux de polynômes,
- une séquence de noeuds  $\tau$ , qui entraîne la connaissance des points de rupture (jointure) des sous-intervalles.

Nous ne définirons pas précisément les B-splines , mais nous nous contenterons de citer les quelques propriétés spécifiques suivantes :

- Nombre de fonctions de base = ordre + nombre de noeuds intérieurs ( $\tau_0$  et  $\tau_L$  exclus).
- Support compact : une fonction B-spline de base d’ordre  $m$  est non-nulle et positive sur au plus  $m$  sous-intervalles adjacents (ceci a des conséquences sur le calcul des coefficients d’une fonction spline dans la base des B-splines : la matrice des produits scalaires des fonctions de base entre elles est une matrice bande, avec au plus  $m$  diagonales non nulles).
- Forme : la forme des B-splines est définie par les noeuds. Pour des noeuds équidistants par exemple, toutes les fonctions de base ont la même forme.
- Baisse de continuité aux extrémités de l’intervalle  $T$  : quand on veut approcher une fonction sur un intervalle  $T$ , on ne connaît généralement pas son comportement en dehors de l’intervalle et il peut être naturel de vouloir des fonctions moins régulières sur les bords, ceci est réalisable dans le cas des B-splines en choisissant une série de noeuds avec plusieurs noeuds égaux sur les bords.
- Somme des valeurs des fonctions de base B-spline en tout point  $t$  égale à 1.
- Base de Riesz : la famille des B-splines n’est pas orthonormée, mais elle a des propriétés qui font qu’elles se rapprochent d’une telle base.

### 3. Bases des Ondelettes :

Les ondelettes combinent les avantages de la base de Fourier (approximation "en fréquence" des fonctions) et des splines (bases localisées). Une base d’ondelette est une famille totale dans  $L^2(\mathbb{R})$ , de fonctions non nécessairement orthonormées, mais formant une base de Riesz. Les fonctions de base sont définies par dilatation-translation d’une fonction  $\psi$  dite mère des ondelettes , de la forme suivante :

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), t \in \mathbb{R}.$$

Le développement d'une fonction dans la base associée donne une analyse multirésolution au sens où le coefficient du développement d'indices  $j, k$  apporte des informations sur la fonction au voisinage de la position  $2^{-jk}$  (localisation) à l'échelle  $2^{-j}$  (fréquence proche de  $2^j$ ).

Grâce à la localisation et à des résultats mathématiques poussés, on peut espérer approcher finement une fonction par un développement dans une base d'ondelettes avec seulement un petit nombre de coefficients non nuls à calculer. (Mallat 2009)

# Données fonctionnelles et méthodes de lissage

## 2.1 Introduction

Nous avons vu que toute analyse de données considérées comme fonctionnelles devait se donner comme première étape la reconstitution des données sous leur forme continue, fonctionnelle justement, à partir des données brutes qui sont toujours discrétisées. Supposons qu'un échantillon  $n$  de données fonctionnelles est collecté sous forme discrétisée, avec pour tout  $i \in \{1, \dots, n\}$ ,  $y_i = (y_{i,1}, \dots, y_{i,p_i})$ . La première tâche du statisticien est donc de convertir ces mesures en une fonction  $x_i$ , dont les valeurs  $x_i(t)$  peuvent être calculées pour toute valeur de  $t$  dans un certain ensemble.

On distingue deux cas :

1. Les observations sont obtenues sans erreur (ou les erreurs de mesures sont négligeables) : dans ce cas, les égalités  $y_{i,j} = x_i(t_{i,j}), j = 1, \dots, p_i$  résument l'observation de la courbe  $x_i$ . Pour trouver  $x_i$ , on fait de l'interpolation des points de coordonnées  $(t_{i,j}, y_{i,j})$ .
2. Les observations sont bruitées, ce que l'on modélise de la façon suivante,

$$y_{i,j} = x_i(t_{i,j}) + \epsilon_{i,j}, \quad j = 1, \dots, p_i, \quad i = 1, \dots, n,$$

où les  $\epsilon_{i,j}$  sont des variables non observées centrées, admettant une variance (inconnue), représentant les erreurs de mesure : c'est un bruit, une perturbation qui contribue au caractère brut des données. Il faut alors faire du lissage, pour ôter l'erreur de mesure, c'est-à-dire prendre en compte le bruit qui se superpose au signal, de le filtrer.

On se concentrera sur le second cas, le plus fréquent, dans ce chapitre. On simplifie ici les notations, en supposant que l'on dispose d'une seule trajectoire : on part de  $y = (y_1, \dots, y_p) \in \mathbb{R}^p$  pour reconstruire une fonction  $x$ . La prise en compte du bruit se superposant au signal se traduit par le modèle de régression à design fixe suivant :

$$y_j = x(t_j) + \epsilon_j, \quad j = 1, \dots, p,$$

où les  $\epsilon_j$  sont des variables non observées i.i.d. centrées, admettant une variance (inconnue). La fonction  $x$ , supposée exister (c'est le point de départ de l'analyse de données fonctionnelles!), est inconnue, on en cherche donc une

approximation ou estimation que l'on notera  $\hat{x}$ , qui ne dépend que de ce qui est connu, les  $(y_j, t_j)_{1 \leq j \leq p}$ .

**Remarque :** Dans le cas où plusieurs trajectoires sont disponibles, deux stratégies principales peuvent être adoptées : estimer chaque courbe pour chaque individu si les observations sont suffisamment denses, ou regrouper les observations pour inférer des structures de population comme les fonctions de moyenne et de covariance. Nous nous concentrons ici uniquement sur la première stratégie.

On distingue plusieurs exigences pour la construction de  $\hat{x}$ .

1. On veut tout d'abord naturellement que la courbe reconstruite se rapproche, en un certain sens, des données initiales. Le "sens" le plus communément donné à une certaine proximité entre observation et reconstruction est celui des moindres carrés : on veut rendre petite la somme des carrés des erreurs, c'est-à-dire la somme des carrés des écarts entre observations et reconstruction, c'est le lissage par moindre carré ordinaire, de cette forme la :

$$\sum_{j=1}^p (y_j - x(t_j))^2.$$

2. On peut vouloir aussi imposer une certaine régularité à la courbe que l'on reconstruit, ce qui revient à supposer que  $\hat{x}$  appartient à un espace de fonctions régulières ("lisses"). C'est le cas lorsque l'on cherche à reconstruire non pas seulement  $x$  mais aussi ses dérivées successives. On verra que ceci est relativement contradictoire avec le fait de rendre le critère des moindres carrés petit, d'où la nécessité d'un compromis. Les techniques utilisées pour la recherche de fonctions régulières consistent souvent en la minimisation d'un critère faisant intervenir les normes  $L^2$  des dérivées successives. On cherchera alors à minimiser

$$\sum_{j=1}^p (y_j - x(t_j))^2 + \lambda \int_T (x^{(k)}(t))^2 dt,$$

pour un certain entier  $k$  et une certaine constante  $\lambda$ . On parlera de lissage par moindres carrés pénalisés. (Farraty-Vieu 2006)

## 2.2 Lissage par Moindre Carrés

### 2.2.1 Moindre carrés ordinaire :

**Définition 7** (Ramsey-Silverman 2005) On appelle critère des moindres carrés pour le problème de régression, consistant à ajuster une courbe  $\hat{x}$  aux données  $(t_j, y_j)_{j=1, \dots, p}$  la fonction de contraste suivante :

$$\text{Crit}_{LS} \text{Crit}_{LS} \text{Crit}_{LS} \text{Crit}_{LS}(x) = \sum_{j=1}^p (y_j - x(t_j))^2.$$

Le principe de la méthode des moindres carrés ordinaires est le suivant : on cherche

$$\hat{x} \in \arg \min_x \text{Crit}_{LS} \text{Crit}_{LS} \text{Crit}_{LS} \text{Crit}_{LS}(x).$$

La question est de savoir sur quel espace de fonctions l'arg min est déterminé, c'est-à-dire sous quelle forme on cherche  $x$  (à quel espace de fonction appartient-elle?). Tenant compte de ce qui précédaient au Chapitre 1, la méthode la plus classique consiste à chercher  $\hat{x}$  sous la forme d'une combinaison linéaire de fonctions de base données. On cherche

donc le minimum du critère ci-dessus sur le sous-espace

$$S_D = \text{Vect}\{\varphi_1, \dots, \varphi_D\}$$

pour  $(\varphi_k)_{k=1, \dots, D}$  des fonctions linéairement indépendantes de  $L^2(T)$ . Cet espace est appelé espace d'approximation, ou encore modèle ("sieve"). Or,

$$\min_{x \in S_D} \text{Crit}_{LS} \text{Crit}_{LS} \text{Crit}_{LS} \text{Crit}_{LS}(x) = \min_{x \in S_D} \sum_{j=1}^p (y_j - x(t_j))^2 = \min_{\theta \in \mathbb{R}^D} \sum_{j=1}^p \left( y_j - \sum_{k=1}^D \theta_k \varphi_k(t_j) \right)^2 = \min_{\theta \in \mathbb{R}^D}$$

Ainsi, la solution à ce problème  $\hat{x}$  aura la forme suivante :

$$\hat{x}(t) = \sum_{k=1}^D \hat{\theta}_k \varphi_k(t), \quad t \in T, \text{ avec } \hat{\theta} = (\hat{\theta}_k)_{k \in \{1, \dots, D\}} \in \arg \min_{\theta \in \mathbb{R}^D} \text{Crit}_{LS, \theta} \text{Crit}_{LS, \theta} \text{Crit}_{LS, \theta} \text{Crit}_{LS, \theta}(\theta).$$

En notant, comme au chapitre précédent,  $\Phi(t) = (\varphi_1(t), \dots, \varphi_D(t))^t$  le vecteur dont les éléments sont les  $D$  premiers éléments de la base évalués en  $t$ , on a la représentation matricielle suivante :

$$\hat{x}(t) = \Phi(t)^t \hat{\theta}.$$

La solution à ce problème d'optimisation est la suivante.

**Proposition 3** (Hastie 2009) Avec les notations ci-dessus, la solution  $\hat{\theta}$  au problème d'optimisation satisfait :

$$\hat{\theta} = (\Phi^t \Phi)^{-1} \Phi^t \mathbf{y},$$

où  $\mathbf{y} = (y_1, \dots, y_p)^t$  est le vecteur des observations et  $\Phi$  la matrice à  $p$  lignes et  $D$  colonnes définie par :

$$\Phi = (\varphi_k(t_j))_{1 \leq j \leq p, 1 \leq k \leq D}$$

Le vecteur  $\hat{y} = (\hat{y}_1, \dots, \hat{y}_p)^t = (\hat{x}(t_1), \dots, \hat{x}(t_p))$  des valeurs ajustées est donc

$$\hat{y} = \Phi \hat{\theta} = \Phi (\Phi^t \Phi)^{-1} \Phi^t \mathbf{y}.$$

**Remarque :** pour des variance qui varient en fonction des temps d'observation on utilise généralement une alternative aux moindres carrés ordinaires, les moindres carrés pondérés. Si on introduit  $w_1, \dots, w_p$  des poids positifs, le critère est :

$$\hat{x} \in \arg \min_x \sum_{j=1}^p w_j (y_j - x(t_j))^2$$

Et, en définissant  $\mathbf{W} = \text{diag}(w_1, \dots, w_p)$  comme la matrice diagonale des poids, l'analogue de la Proposition 3 est :

$$\hat{\theta} = (\Phi^t \mathbf{W} \Phi)^{-1} \Phi^t \mathbf{W} \mathbf{y}.$$

## 2.2.2 Choix de la dimension de l'espace d'approximation

Les méthodes précédentes posent la question du choix, crucial, de la dimension  $D$  de l'espace  $S_D$  d'approximation, c'est-à-dire celui du nombre de coefficients  $\hat{\theta}_k$  estimés.

On peut constater en pratique la nécessité de faire un compromis entre de petites valeurs pour  $D$  et de grandes valeurs. Ceci peut aussi être démontré en théorie, en calculant un risque d'estimation de  $x$  par  $\hat{x}$ , risque quadratique intégré par exemple, et en cherchant à le rendre le plus petit possible. L'étude de ce risque montre qu'un compromis biais-variance est requis :

- (i) Si  $D$  est choisi trop petit, le biais de l'estimation va être trop grand. On a trop peu de flexibilité, les données sont sous-ajustées ("underfitting").
- (ii) Inversement, si  $D$  est choisi grand, on a beaucoup de flexibilité mais ceci peut mener à un sur-ajustement des données ("overfitting").

Le problème du choix de  $D$  a été largement étudié dans la littérature, dans le cadre de problèmes non-paramétriques généraux. Des méthodes variées ont été proposées, comme la validation croisée ou la sélection de modèle par pénalisation de contraste.(voir par exemple Massart 2007)

Par ailleurs, il est utile de garder en mémoire qu'augmenter  $D$  ne se traduit pas toujours de la même façon, en fonction du type de fonctions de base choisie. Dans le cas de la base de Fourier, c'est simple, mais dans le cas d'une base de splines, c'est plus compliqué.

Dans le cas du lissage de données fonctionnelles, pour parer ces inconvénients des moindres carrés ordinaires, la méthode la plus fréquemment utilisée pour effectuer de manière automatique le compromis biais-variance, en tenant compte de la régularité des courbes est l'utilisation de splines de lissage, dans le cadre du lissage par moindres carrés pénalisés

## 2.3 Lissage par moindres carrés pénalisés :

### 2.3.1 Principe

L'approche par moindres carrés pénalisés vise à produire un estimateur  $\hat{x}$  pour une courbe  $x$  en minimisant un critère qui rend explicite deux objectifs contradictoires en estimation de courbe :

- on veut bien sûr que la courbe estimée donne un bon ajustement aux données,
- mais que l'ajustement ne soit pas trop précis, sous peine d'obtenir une courbe excessivement oscillante (sur-ajustement).

Il va donc s'agir de minimiser un critère des moindres carrés, mais sous la contrainte que la fonction cible appartient à un espace de fonctions "lisses". Il faut alors choisir un tel espace, parmi une large gamme d'espace de fonctions régulières, et déterminer un critère permettant de mesurer la "rugosité" (roughness) d'une fonction.

Le critère classiquement utilisé en analyse de données fonctionnelles est celui de la convergence des integrales suivantes

$$\text{pen}_m(x) = \int_T (x^{(m)}(t))^2 dt$$

Ainsi,  $x^{(1)} = x'$  mesure par exemple la pente de  $x$  à chaque instant,  $x^{(2)} = x''$  mesure la courbure de  $x$  en chaque point. Prendre l'intégrale du carré des fonctions correspondantes revient à obtenir une mesure globale, souvent appelée énergie.

En tenant compte des objectifs contradictoires ci-dessus, on va donc cette fois chercher à minimiser le critère suivant :

**Définition 8** On appelle critère des moindres carrés pénalisés, avec pénalité de lissage pour le problème de régression, consistant à ajuster une courbe  $x$  aux données  $(t_j, y_j)_{j=1, \dots, p}$ , le critère suivant :

$$\text{CritPenLS} \text{ CritPenLS} \text{ CritPenLS} \text{ CritPenLS}(x) = \sum_{j=1}^p (y_j - x(t_j))^2 + \lambda \text{pen}_m(x) \text{ où } \text{pen}_m(x) = \int_T [x^{(m)}(t)]^2 dt \text{ où } \lambda > 0 \text{ est un paramètre à calibrer, appelé paramètre de lissage.}$$

Le principe est donc de chercher  $\hat{x} \in \arg \min_x$

**Théorème 2** On pose :

$$\hat{x} \in \arg \min_{x \in \mathcal{S}^m(T)} \text{CritPenLS}(x)$$

Alors  $\hat{x}$  est unique, et est appelée spline naturelle de lissage d'ordre  $m$  associée aux données  $(t_j, y_j)_{j=1, \dots, p}$ . Elle est définie sur  $T = [a, b]$  de la façon suivante :

- i) la restriction de  $\hat{x}$  aux intervalles  $[a, t_1]$  et  $[t_p, b]$  est un polynôme de degré au plus  $m - 1$  ;
- ii) la restriction de  $\hat{x}$  aux intervalles  $[t_j, t_{j+1}]$ ,  $j = 1, \dots, p - 1$ , est un polynôme de degré au plus  $2m - 1$  ;
- iii) la fonction  $\hat{x}$  est de classe  $\mathcal{C}^{2m-2}(T)$ .

En particulier,  $\hat{x}$  est une spline dont les nœuds sont les points  $t_j$  où il y a des données (Boor 2001).

D'autres types de pénalités sont considérables, notamment dans le cas de données périodiques (voir Massart 2007)

### 2.3.2 Choix du paramètre de lissage

Pour une simplicité d'explication, on se place dans le cas  $m=2$  (le plus utilisé en pratique)

Le paramètre de lissage  $\lambda$  apparaissant devant la pénalité dans le critère CritPenLS de la Définition 8 peut être vu comme le multiplicateur de Lagrange apparaissant dans tout problème d'optimisation sous contrainte (théorème des extrema liés) - ici minimisation du contraste des moindres carrés sous contrainte de régularité. Par conséquent, il est inévitable.

En pratique, la calibration de  $\lambda$  doit tenir compte des remarques suivantes :

- Si  $\lambda$  est petit, le poids de la pénalité dans la minimisation du critère est peu important, on privilégie l'ajustement aux données. Dans le cas extrême où  $\lambda = 0$ , minimiser le critère sur  $\mathcal{S}^m(T)$  revient à interpoler les données, sans tenir compte de la régularité de la courbe.
- Si  $\lambda$  est grand, le caractère lisse de la fonction choisie sera privilégié par rapport à la fidélité aux données. Dans



le cas extrême où  $\lambda = \infty$ , la solution au problème de minimisation sera la droite de régression des  $y_j$  sur les  $t_j$ .

La méthode la plus souvent utilisée est la validation croisée leave-one-out. Pour tout  $j = 1, \dots, p$ , on retire l'observation  $(t_j, y_j)$  de l'échantillon de données, et on calcule la solution au problème de minimisation défini à partir des autres données :

$$\hat{x}_\lambda^{(-j)} \in \arg \min_{x \in \mathcal{S}^2(T)} \sum_{j' \neq j} (y_{j'} - x(t_{j'}))^2 + \lambda \int_T (x''(t))^2 dt$$

La spline obtenue  $\hat{x}_\lambda^{(-j)}$  dépend du paramètre  $\lambda$ . On voit alors  $(t_j, y_j)$  comme une nouvelle observation, et on considère l'erreur de prévision : qualité de  $\hat{x}_\lambda^{(-j)}(t_j)$  pour estimer  $y_j$ , et ceci pour tout indice  $j$ . Ceci nous conduit au critère suivant.

**Définition 9** On appelle score de validation croisée pour le problème de régression à partir des données  $(t_j, y_j)_{j=1, \dots, p}$  le critère suivant, avec les notations introduites ci-dessus, et pour tout  $\lambda \geq 0$  :

$$CV(\lambda) = \frac{1}{p} \sum_{j=1}^p \left( y_j - \hat{x}_\lambda^{(-j)}(t_j) \right)^2$$

Cette méthode de validation croisée permet de déterminer de manière automatique le paramètre de lissage  $\lambda$  optimal, en minimisant le score  $CV(\lambda)$  (Hastie 2009).

Critère servant à prouver l'existence et l'unicité d'une spline minimisante (voir Reinsch 1964)

# Estimation non paramétrique de la densité de probabilité par la méthode des fonctions orthogonales

**Proposition 4** (Conway 1990) *L'espace  $L^2$  muni du produit scalaire  $(f, g) = \int_{\mathbb{R}} f(t)g(t)dt$  est un espace de Hilbert.*

**Définition 10** (Stark 1986) *Un ensemble dénombrable  $\mathcal{B} = \{e_n, n \in \mathbb{N}\}$  dans  $L^2$  est orthonormé si :*

$$(e_n, e_m) = \begin{cases} 1 & \text{si } n = m \\ 0 & \text{sinon} \end{cases}$$

*Cela signifie que les vecteurs  $e_n$  forment une base orthonormée de l'espace  $L^2$ .*

## 3.1 Description de la méthode

Soit  $(X_n, n \in \mathbb{N})$  une suite de variables aléatoires indépendantes et identiquement distribuées définies sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  de densité de probabilité inconnue  $f$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ .

Il s'agit alors d'estimer  $f$  à partir de l'échantillon  $X_1, \dots, X_n$ . Pour cela, on suppose que :

- L'espace de Hilbert  $L^2$  est de dimension infinie muni de sa norme usuelle  $\|\cdot\|$  de son produit scalaire  $(\cdot, \cdot)$ ;
- $\{e_k, k \in \mathbb{N}\}$  un système orthonormé dans  $L^2$ ;
- $f \in L^2$  tel que :

$$f(t) = \sum_{k=0}^{\infty} a_k e_k(t), \quad t \in \mathbb{R};$$

- Le développement à l'ordre  $d$ , de  $f(t)$  est :

$$f_{d_n}(t) = \sum_{k=0}^{d_n} a_k e_k(t), \quad t \in \mathbb{R}.$$

Avec  $a_k = \int e_k(x) \cdot f(x) dx = \mathbb{E}[e_k(X)]$ .

Par la méthode des moments, on peut estimer  $a_k$  par :

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^n e_k(X_i).$$

Ainsi,  $f(t)$  peut être estimée par :

$$\hat{f}_{d_n}(t) = \sum_{k=0}^{d_n} \hat{a}_k e_k(t) = \sum_{k=0}^{d_n} \frac{1}{n} \sum_{i=1}^n e_k(X_i) e_k(t) = \frac{1}{n} \sum_{i=1}^n K_{d_n}(t, X_i).$$

ou

$$K_{d_n}(t, X) = \sum_{k=0}^{d_n} e_k(X) e_k(t).$$

Le terme  $K_{d_n}(t, X)$  désigne le noyau de l'estimateur. L'expression de  $\hat{f}_{d_n}(t)$  révèle que cet estimateur, par la méthode des fonctions orthogonales, appartient à la classe des estimateurs à noyau. Ainsi, il peut être représenté comme une somme pondérée de variables aléatoires indépendantes. Par conséquent, l'analyse de l'estimateur  $f_{d_n}(t)$  peut être abordée à partir des théorèmes généraux régissant les estimateurs à noyau.

### 3.1.1 Propriétés de l'estimateur

Les bases  $\hat{a}_k$  sont des estimateurs sans biais de  $a_k$  (Stoica 2005). En effet,

$$\mathbb{E}[\hat{a}_k] = \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n e_k(X_i) \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[e_k(X_i)] = \mathbb{E}[e_k(X)]$$

$\hat{f}_{d_n}(t)$  est un estimateur sans biais de  $f_{d_n}(t)$  (popoulis 2002). En effet,

$$\mathbb{E}[\hat{f}_{d_n}(t)] = \mathbb{E} \left[ \sum_{k=0}^{d_n} \hat{a}_k e_k(t) \right] = \sum_{k=0}^{d_n} \mathbb{E}[\hat{a}_k] e_k(t) = \sum_{k=0}^{d_n} a_k e_k(t) = f_{d_n}(t)$$

Le biais de  $\hat{f}_{d_n}(t)$  est donné par (Stark 1986) :

$$\text{Biais}[\hat{f}_{d_n}(t)] = \mathbb{E}[\hat{f}_{d_n}(t)] - f(t) = \sum_{k=0}^{d_n} a_k e_k(t) - f(t) = - \sum_{k=d_{n+1}}^{\infty} a_k e_k(t) = f_d(t) - f(t)$$

Cette relation montre que  $\hat{f}_{d_n}(t)$  est un estimateur biaisé de  $f(t)$ .

Nous avons présenté la méthode d'estimation de densité par fonctions orthogonales, ainsi que les propriétés des

estimateurs correspondants. Le prochain point sera d'étudier ces estimateurs pour des bases orthogonales spécifiques comme Fourier, splines ou ondelettes. Le choix de la base est crucial pour la qualité de l'estimation.

## 3.2 Estimateurs associés aux fonctions trigonométriques

### 3.2.1 Estimateur de Dirichlet

Nous considérons l'intervalle  $I = [-\pi, \pi]$  comme un sous-ensemble de  $\mathbb{R}$  muni de la mesure de Lebesgue. La base orthonormée choisie est :

$$e_0 = \frac{1}{\sqrt{2\pi}}; e_{2k}(t) = \frac{\cos(kt)}{\sqrt{\pi}}; e_{2k+1}(t) = \frac{\sin(kt)}{\sqrt{\pi}} \text{ pour } t \in [-\pi, \pi], k \geq 1$$

Le noyau de Dirichlet associé est alors (Conway 1990) :

$$K_{d_n}(x, t) = \frac{1}{2\pi} \frac{\sin \left[ \frac{d_n(x-t)}{2} \right]}{\sin \left[ \frac{x-t}{2} \right]} \text{ pour } x \neq t$$

L'estimateur de  $f(t)$  est alors de la forme :

$$\hat{f}_{d_n}(t) = \frac{1}{2\pi n} \sum_{i=1}^n \frac{\sin \left[ \frac{d_n(X_i-t)}{2} \right]}{\sin \left[ \frac{X_i-t}{2} \right]}$$

### 3.2.2 Propriété

#### BIAIS :

le biais de l'estimateur de Dirichlet converge vers 0 lorsque le nombre de termes  $d_n$  augmente, avec un taux de convergence en  $\mathcal{O}(1/\sqrt{d_n})$ .

En effet :

$$B(\hat{f}_{d_n}(t)) = \mathbb{E} \left[ \sum_{k=0}^{d_n} \hat{a}_k e_k(t) \right] - \sum_{k=-\infty}^{+\infty} a_k e_k(t) = \sum_{k=0}^{d_n} \mathbb{E}[\hat{a}_k] e_k(t) - \sum_{k=-\infty}^{+\infty} a_k e_k(t).$$

Or, on sait que  $\mathbb{E}[\hat{a}_k] = a_k + \mathcal{O}(1/\sqrt{n})$ , où  $n$  est la taille de l'échantillon.

Donc, le biais s'écrit :

$$B(\hat{f}_{d_n}(t)) = \sum_{k=0}^{d_n} \left( a_k + \mathcal{O} \left( \frac{1}{\sqrt{n}} \right) \right) e_k(t) - \sum_{k=-\infty}^{+\infty} a_k e_k(t) = \sum_{k=0}^{d_n} \mathcal{O} \left( \frac{1}{\sqrt{n}} \right) e_k(t).$$

En utilisant le fait que la série  $\sum_{k=0}^{d_n} e_k(t)$  converge vers  $f(t)$  quand  $d_n$  tend vers l'infini, on en déduit que le biais est de l'ordre de  $\mathcal{O}(1/\sqrt{d_n})$  (Stark 1986).

**Variance :**

La variance de l'estimateur de Dirichlet converge vers 0 lorsque le nombre de termes  $d_n$  augmente, avec un taux de convergence en  $\mathcal{O}(1/d_n)$ .

En effet,

La variance de l'estimateur  $f_{d_n}^\wedge(t)$  est définie par :

$$V((t)) = E[(t) - E[(t)]]^2$$

En développant, on obtient :

$$V(f_{d_n}^\wedge(t)) = \mathbb{E} \left[ \left( \sum_{k=0}^{d_n} (\hat{a}_k - \mathbb{E}[\hat{a}_k]) e_k(t) \right)^2 \right] = \sum_{k=0}^{d_n} \text{Var}(\hat{a}_k) |e_k(t)|^2.$$

Or, on sait que  $\text{Var}(\hat{a}_k) = \mathcal{O}(1/n)$ , où  $n$  est la taille de l'échantillon.

Donc, la variance s'écrit :

$$V(f_{d_n}^\wedge(t)) = \sum_{k=0}^{d_n} \mathcal{O} \left( \frac{1}{n} \right) \cdot 1 = \mathcal{O} \left( \frac{d_n}{n} \right).$$

En utilisant le fait que  $d_n = \mathcal{O}(n)$ , on en déduit que la variance est de l'ordre de  $\mathcal{O}(1/d_n)$  (Stoica 2005).

**Erreur quadratique moyenne intégrée :**

Le MISE de l'estimateur de Dirichlet s'écrit :

$$\text{MISE}() = \frac{1}{n} \int \frac{f(t)}{p(t)} dt$$

En effet,

$$\text{MISE}() = E \left[ \int (f_{d_n}^\wedge(t) - f(t))^2 dt \right] = \int E \left[ (f_{d_n}^\wedge(t) - f(t))^2 \right] dt$$

Or, pour l'estimateur de Dirichlet, on a :

$$\mathbb{E} [f_{d_n}^\wedge(t)] = \frac{1}{n} \sum_{i=1}^n \frac{1}{p(X_i)} \mathbb{I}_{X_i \leq t} = \frac{1}{n} \int_{-\infty}^t \frac{1}{p(u)} dF(u).$$

Où  $F(t)$  est la fonction de répartition de la densité  $f(t)$ . En utilisant cette expression, on obtient :

$$\begin{aligned} E \left[ (f_{d_n}^\wedge(t) - f(t))^2 \right] &= E \left[ \left( \frac{1}{n} \int_{-\infty}^t \frac{1}{p(u)} dF(u) - f(t) \right)^2 \right] = \frac{1}{n^2} \int_{-\infty}^t \frac{f(u)}{p(u)} du - 2 \frac{1}{n} \frac{f(t)}{p(t)} + f(t)^2 \\ &= \frac{1}{n} \frac{f(t)}{p(t)} - f(t)^2 \end{aligned}$$

En intégrant cette expression sur  $t$ , on obtient finalement :

$$\text{MISE}() = \int \left( \frac{1}{n} \frac{f(t)}{p(t)} - f(t)^2 \right) dt = \frac{1}{n} \int \frac{f(t)}{p(t)} dt$$

Ce résultat montre que le MISE de l'estimateur de Dirichlet dépend de la densité  $f(t)$  et de la distribution de Dirichlet  $p(t)$  utilisée, et qu'il décroît en  $1/n$  lorsque la taille de l'échantillon augmente (Popoulis 2002).

Nous citerons l'estimateur associé à la base de Fejer et de Legendre et bien d'autre comme des estimateur adapter à une fonction presentant un caractère périodique.

**Remarque :** Les splines étant largement mentionnées dans la literature , nous ommetrons leur présentation en citant les splines cubiques , les B-spline (voir Boor 2001) ainsi que les splines de lissage (voir Rendall 1999) comme les principaux estimateurs relative à cette base la .

### 3.3 Estimateur associé a la base des Ondelette

#### 3.3.1 Estimateur de densité de Haar

L'estimateur de Haar pour la densité de probabilité, noté  $\hat{h}(x)$ , est défini comme la somme pondérée des fonctions de base de Haar décalées et dilatées, chacune multipliée par un coefficient estimé à partir des données :

$$\hat{h}(x) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k} \phi_{j,k}(x)$$

où  $\phi_{j,k}$  est la fonction de base de Haar décalée et dilatée, et  $d_{j,k}$  sont les coefficients de détail estimés (Mallat 2008).

#### 3.3.2 Propriétés

**Biais :**

Le biais de l'estimateur de Haar est nul :

$$\text{Biais}(\hat{h}(x)) = \mathbb{E}[\hat{h}(x)] - h(x) = 0$$

En effet

on peut écrire la fonction de régression  $h(x)$  sur la base orthogonale de Haar :

$$h(x) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \theta_{j,k} \phi_{j,k}(x)$$

Où  $\phi_{j,k}(x)$  sont les fonctions de base de Haar, définies par :  $\phi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \cdot \mathbf{1}\{x \in I_{j,k}\}$

Et  $\theta_{j,k}$  sont les coefficients de la décomposition de  $h(x)$  sur cette base.

Alors, l'estimateur de Haar s'écrit :

$$\hat{h}(x) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \frac{1}{\sqrt{2^j}} \cdot \frac{1}{\sqrt{n}} \cdot X_k \cdot \mathbf{1}\{x \in I_{j,k}\}$$

Où  $X_k$  sont les observations dans les sous-intervalles  $I_{j,k}$ .

Calcul de l'espérance :

$$\mathbb{E}[\hat{h}(x)] = \mathbb{E} \left[ \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \frac{1}{\sqrt{2^j}} \frac{1}{\sqrt{n}} X_k \mathbf{1}\{x \in I_{j,k}\} \right] = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \frac{1}{\sqrt{2^j}} \frac{1}{\sqrt{n}} \mathbb{E}[X_k] \mathbf{1}\{x \in I_{j,k}\}$$

$$= \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \frac{1}{\sqrt{2^j}} \frac{1}{\sqrt{n}} \theta_{j,k} \phi_{j,k}(x) = h(x)$$

**Variance :** La variance de l'estimateur de Haar est :

$$\text{Var}(\hat{h}(x)) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \frac{\sigma^2}{2^j n} \cdot \mathbf{1}\{x \in I_{j,k}\}$$

En effet : à partir de l'expression de l'estimateur de Haar vu précédemment

$$\begin{aligned} \text{Var}(\hat{h}(x)) &= \text{Var} \left[ \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \frac{1}{\sqrt{2^j}} \frac{1}{\sqrt{n}} X_k \mathbf{1}\{x \in I_{j,k}\} \right] \\ &= \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \frac{1}{2^j n} \text{Var}(X_k) \mathbf{1}\{x \in I_{j,k}\} \\ &= \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \frac{\sigma^2}{2^j n} \mathbf{1}\{x \in I_{j,k}\} \end{aligned}$$

Ainsi, le MISE de l'estimateur de Haar dépend uniquement de la variance des observations et décroît lorsque le nombre d'observations  $n$  augmente (Mallat 2008).

Nous citerons l'ondelette de Debauchis , tres utilisées en analyse multirésolution comme un autre estimateur associé a la base des ondelettes .

# APPLICATION

## 4.1 Introduction

Dans ce chapitre, nous mettrons en lumière l'utilité pratique des techniques d'estimation non paramétrique de la densité en les appliquant à des données météorologiques. L'analyse des données météorologiques est cruciale pour comprendre les tendances climatiques, prévoir les conditions météorologiques et prendre des décisions informées dans divers domaines tels que l'agriculture, l'urbanisme et la gestion des catastrophes naturelles.

Nous explorerons comment les méthodes non paramétriques peuvent être utilisées pour estimer des distributions de variables climatiques telles que la température, les précipitations et la vitesse du vent. À travers des exemples concrets, nous démontrerons comment ces techniques peuvent révéler des motifs cachés dans les données et fournir une compréhension approfondi pour les chercheurs et les décideurs

## 4.2 Logiciel utilisé :

**R** est un logiciel libre et un environnement de développement intégré dédié à la statistique et à la science des données, cree par Ross Ihaka et Robert Gentleman , il a été conçu pour le calcul statistique et la création de graphiques, R est un outil puissant et polyvalent utilisé par les chercheurs, les statisticiens et les data scientists du monde entier.

### 4.2.1 Caractéristiques Principales

- **Langage de Programmation Statistique** : R est conçu spécifiquement pour les tâches statistiques et permet de manipuler, d'analyser et de visualiser des données de manière efficace.
- **Large Éventail de Packages** : Avec des milliers de packages disponibles via le Comprehensive R Archive Network (CRAN), R offre une gamme étendue de méthodes statistiques et graphiques.
- **Visualisation de Données** : R excelle dans la création de graphiques élégants et informatifs grâce à des packages comme ggplot2.
- **Communauté Active** : R bénéficie d'une communauté dynamique qui contribue régulièrement à son amélioration, offrant un soutien précieux sous forme de documentation, de forums et de tutoriels.
- **Intégration avec d'Autres Outils** : R peut s'intégrer avec d'autres logiciels et langages de programmation, facilitant le flux de travail et l'analyse de données à grande échelle.



## 4.2.2 Utilisation de R pour l'Estimation Non Paramétrique

Dans le contexte de l'estimation non paramétrique de la densité, R offre plusieurs packages puissants qui facilitent l'application de ces techniques. Pour des données fonctionnelles, le package **fda** (Functional Data Analysis) est particulièrement utile. Les données fonctionnelles sont des données où chaque observation est une fonction, typiquement dépendant du temps ou de l'espace, comme les courbes de température ou les profils de vent.

## 4.3 Données Utilisées

Les données utilisées dans cette étude sont des données météorologiques couvrant la période de février 2009 à février 2019. Elles proviennent de la majorité des wilayas d'Algérie. Ces données ont été extraites d'un historique disponible sur le site GitHub.com.

Les données météorologiques ont été collectées de février 2009 à février 2019, fournissant ainsi une très large couverture pour les villes incluses (notamment la ville de Bejaia qui sera la cible de notre application). Ces enregistrements sont accessibles sur GitHub (AbdouTlili 2019), offrant une source fiable et facilement exploitable pour notre analyse.

Les données originales comprennent plusieurs paramètres climatiques, notamment la date d'observation, la température enregistrée, le niveau de précipitation, la direction du vent et la pression atmosphérique.

Pour les besoins de notre application, nous nous concentrons exclusivement sur les températures indexées par le temps. Cette variable présente les caractéristiques nécessaires pour notre étude d'estimation non paramétrique de la densité. En nous limitant à l'analyse des températures, nous pouvons mieux explorer les tendances saisonnières et les variations climatiques dans les villes étudiées.

Les détails du code R utilisé pour produire la suite de ce chapitre peuvent être retrouvés dans l'ouvrage de Ramsey-Silverman (2009).

## 4.4 Algorithme

La simulation que nous allons utiliser comporte les étapes suivantes

- Calcul des températures moyennes pour la ville de Bejaia et représentation graphique des données.
- Analyse et formulation d'hypothèses sur la nature des données.
- Calcul de la densité vraie des données observées, représentation graphique et interprétation.
- Estimation de la densité avec des fonctions orthogonales à base Fourier et représentation graphique superposée à la densité vraie pour une meilleure visualisation.
- Calcul du MISE de l'estimateur et interprétation.

### 4.4.1 Calcul des moyennes et représentation

Nous avons calculé les moyennes mensuelles des températures, ce qui nous a permis de dégager des variations saisonnières et d'identifier des anomalies climatiques éventuelles. Les résultats obtenus sont ensuite représentés graphiquement ci-dessous afin de fournir une visualisation claire et intuitive des données. Le graphique montre l'évolution des températures au cours des années, mettant en évidence les tendances à long terme.

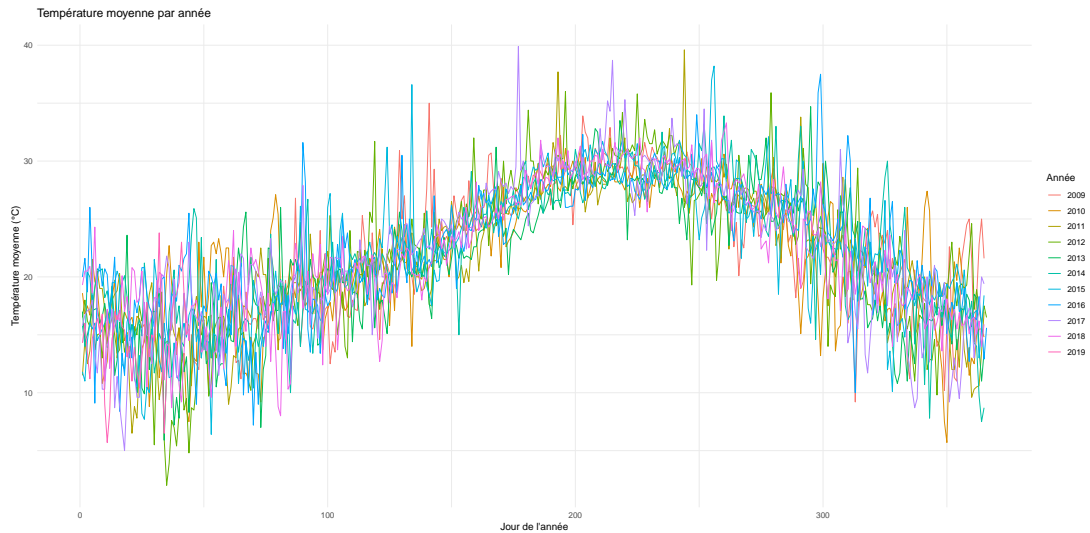


FIGURE 4.1 – Courbes des températures moyennes dans la ville de Bejaia entre 2009 et 2019

#### 4.4.2 Hypothese sur la nature des données

Après avoir visualisé les courbes des températures moyennes à Béjaïa , il apparaît clairement que les données présentent des tendances et un certain caractère périodique. Cette observation nous conduit à choisir la base de Fourier comme étant la base adéquate pour l'estimation de la densité des données. Les séries temporelles de températures montrent des variations saisonnières régulières, rendant la base de Fourier particulièrement adaptée pour capturer ces cycles périodiques et fournir une estimation précise de la densité des températures.

#### 4.4.3 Calcule de la densité vrai et représentation

Nous avons calculé la densité vraie de nos données météorologiques en utilisant la fonction `density()` de R. Cette approche s'est avérée très efficace pour obtenir une représentation précise de la distribution des températures enregistrées sur la période étudiée , la figure si dessous illustre cette distribution .

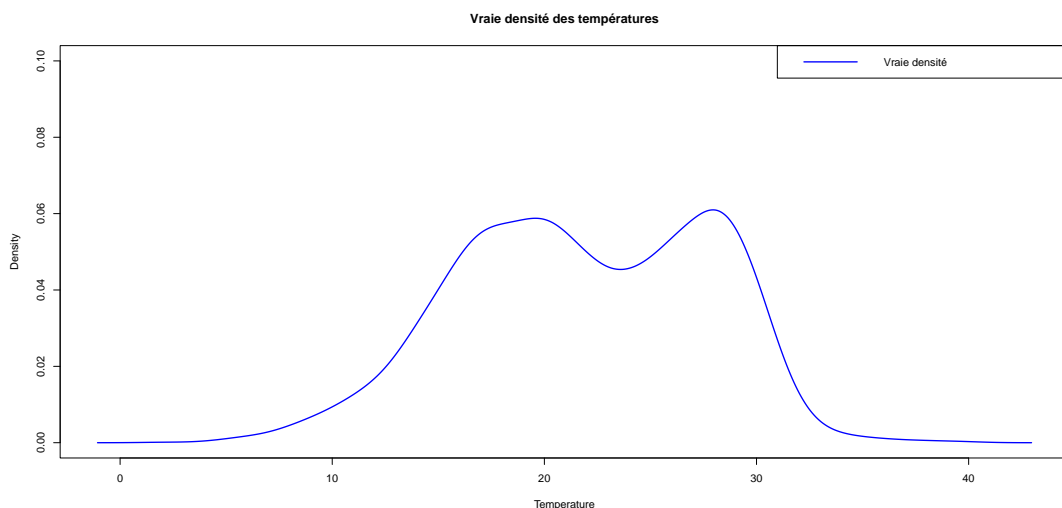


FIGURE 4.2 – Densité réelle des températures à Béjaïa (2009-2019)

**Interprétation :** La courbe de densité des températures à Béjaïa entre 2009 et 2019 présente une distribution gaussienne centrée autour de 20°C, avec une variance estimée à 24.97. Le coefficient d'asymétrie de 0.7896 indique une légère asymétrie positive, suggérant une fréquence légèrement plus élevée de températures supérieures à la moyenne,

typique des températures de la région méditerranéenne .

#### 4.4.4 Estimation de la densité

L'estimation de la densité des températures repose sur l'utilisation de la fonction `density.fd()` du package `Fda`. Avant l'estimation, les données ont été prétraitées : elles ont été transformées du format numérique classique en format fonctionnel, permettant ainsi leur représentation sous forme de courbes continues. Ensuite, un lissage a été appliqué aux vecteurs de températures pour capturer les tendances et les variations saisonnières de manière robuste, la figure 4.3 illustre cette représentation :

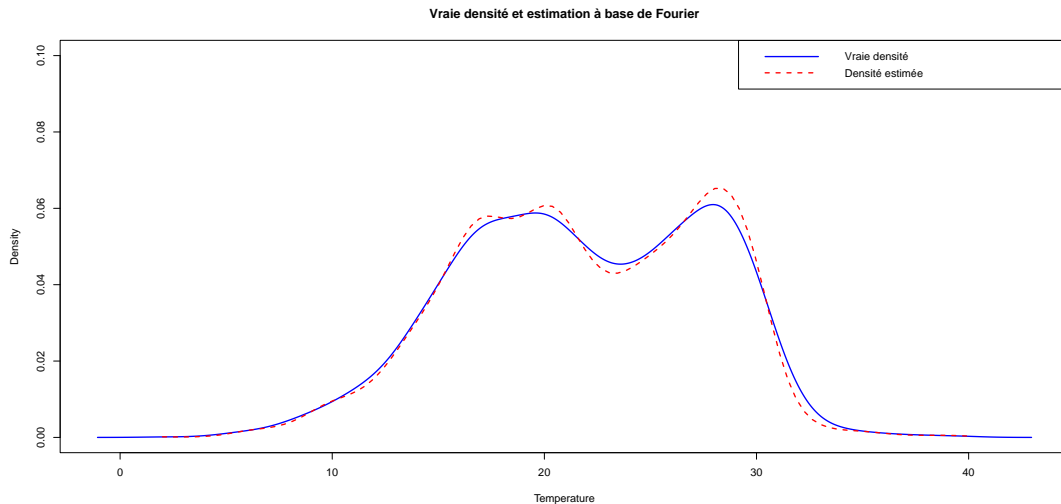


FIGURE 4.3 – Estimation de la Densité à Base de Fourier des Températures a Béjaïa (2009-2019)

#### Interprétation :

A partir de l'observation de la figure 4.3 , nous faisons les constatations suivantes :

- La courbe de la densité estimée est relativement superposée à la courbe de la vraie densité.
- L'erreur d'estimation se localise principalement autour des valeurs modales de la vraie densité.
- une variance estimée à 25.18 , ce qui est assez proche de la variance estimée pour la vraie densité .

L'analyse de la densité des températures à l'aide de l'estimation par fonctions de Fourier révèle une correspondance étroite et cohérente entre la densité estimée et la vraie densité sur l'intervalle de 0 à 40 pour les températures et de 0 à 0.1 pour les densités. Cette similitude suggère que l'estimateur utilisé est de bonne qualité pour capturer la structure sous-jacente des données de température. L'erreur d'estimation, concentrée principalement autour des valeurs moyennes et de valeur supérieure à la moyenne de la vraie densité, renforce cette conclusion en indiquant que les variations les plus significatives des températures sont bien représentées par le modèle. Ainsi, l'estimation par fonctions de Fourier semble appropriée pour modéliser et interpréter la densité des températures pour la ville de Béjaïa sur la période étudiée.

#### 4.4.5 Calcule du MISE :

Nous avons calculé le MISE (Mean Integrated Squared Error) pour les 500 premières observations, et il est égal à  $5.4938 \times 10^{-5}$ . Cette valeur du MISE indique que l'estimation de la densité obtenue est très précise, avec une erreur d'intégration moyenne très faible. Une telle précision est généralement souhaitable, car elle montre que l'estimateur est capable de capturer la structure sous-jacente des données avec un minimum d'erreur.

## 4.5 Conclusion

Dans ce chapitre, nous avons appliqué l'estimation de densité avec la méthode des fonctions orthogonales à des bases de Fourier sur un ensemble de données météorologiques. L'objectif était de comparer la densité estimée à la densité réelle et d'évaluer la qualité de l'estimation.

L'application de cette méthode a montré son potentiel pour fournir des estimations précises et fiables. Cette technique peut être largement utilisée pour analyser des données présentant une périodicité, comme c'est souvent le cas dans les séries temporelles environnementales et climatiques.

En conclusion, à partir d'une hypothèse valide sur la forme de la densité ciblée et d'un ordre de troncature correct sur la série des observations, la méthode des fonctions orthogonales peut être une alternative judicieuse aux autres méthodes d'estimation non paramétrique que sont l'estimation par histogramme et la méthode du noyau de Parzen-Rosenblatt .

# Conclusion générales et perspectives

## Conclusion générales

L'estimation non paramétrique de la densité de probabilité est un domaine fondamental en statistique moderne, offrant une alternative flexible et robuste aux méthodes paramétriques traditionnelles. Ce mémoire a exploré en profondeur les différentes approches non paramétriques, en mettant particulièrement l'accent sur l'utilisation des fonctions orthogonales.

Dans un premier temps, nous avons présenté les concepts théoriques nécessaires à la compréhension des variables aléatoires fonctionnelles, en abordant les espaces de Hilbert et les processus stochastiques. Cette base théorique a permis de comprendre les enjeux et les défis liés à l'estimation non paramétrique dans le contexte des données fonctionnelles. Ensuite, nous avons détaillé plusieurs méthodes d'estimation non paramétrique de la densité en utilisant des fonctions orthogonales, notamment les fonctions trigonométriques, les spline et les ondelettes. Nous avons analysé les propriétés de ces estimateurs, en mettant en lumière leurs avantages et leurs limitations en termes de biais et de variance.

La partie pratique de ce mémoire a illustré l'application de ces méthodes sur des données météorologiques. Cette application concrète a permis de démontrer l'efficacité et la pertinence des approches proposées. En utilisant le logiciel R, nous avons montré comment ces méthodes peuvent être mises en œuvre pour estimer des densités de probabilité à partir de données réelles, tout en discutant des résultats obtenus.

En conclusion, les méthodes d'estimation non paramétrique basées sur les fonctions orthogonales constituent une approche puissante pour analyser des données complexes. Elles offrent une grande flexibilité et permettent de s'affranchir des contraintes imposées par les méthodes paramétriques. Les travaux présentés dans ce mémoire ouvrent la voie à de nombreuses perspectives de recherche.

## Perspectives

Ce travail offre de nombreuses perspectives théorique et pratique :

- D'un point de vue théorique, cette méthode d'estimation peut donner des résultats similaires dans d'autres cas tels que l'estimation des dérivées de la fonction de densité, la fonction de répartition, ou encore, les fonctions caractéristiques.
- un travail tout entier peut être dédié au choix de l'espace d'approximation  $D$  de sorte à trouver le meilleur compromis Biais-Variance
- L'utilisation de la méthode des fonctions orthogonales pour analyser des données sous des hypothèses fortes, comme celles impliquées dans le mouvement brownien, est pertinente.
- la comparaison théorique et numérique par simulation de l'estimateur par la méthode des fonctions orthogonales

avec un estimateur obtenu par la méthode des splines.

# Bibliographie

- [1] Abramowitz, M., Stegun, I. A. (Eds.). (1972). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications. (ISBN : 978-0-486-61272)
- [2] AbdouTlili. (2019). Algerian-Weather-History-2009—2019-. GitHub. <https://github.com/AbdouTlili/Algerian-Weather-History-2009—2019->. Consulté le 01/06/2024.
- [3] Briant, O., Pagès, H. (2000). L'estimation non paramétrique des densités : Méthodes et applications. *Revue de Statistique Appliquée*, 48(3), 5-21.
- [4] Brézis, H. (1983). *Analyse fonctionnelle : Théorie et applications*. Paris : Masson.
- [5] Brézis, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer. ISBN-10 : 3642158303, ISBN-13 : 978-3642158300.
- [6] Brown, A. B. (2010). Splines and their applications. In R. Green (Ed.), *Handbook of Computational Methods* (pp. 89-112). Academic Press.
- [7] Comte, F., Rozenholc, C. (2015). On the adaptivity of block thresholding estimators of the intensity of inhomogeneous Poisson processes. *Electronic Journal of Statistics*, 9(1), 1511-1532.
- [8] Conway, J. B. (1990). *A Course in Functional Analysis*. New York : Springer-Verlag. ISBN 978-0-387-97245-9. OCLC 21195908.
- [9] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics. ISBN : 978-0-89871-274-1.
- [10] de Boor, C. (2001). *A Practical Guide to Splines (Revised Edition)*. Springer. ISBN 978-0387953663.
- [11] Evans, L. C., Gariepy, R. F. (1992). *Measure Theory and Fine Properties of Functions*. CRC Press. 1st edition. ISBN-10 : 0849371570, ISBN-13 : 978-0849371578.
- [12] Farin, G. (2002). *Curves and Surfaces for CAGD : A Practical Guide*. Morgan Kaufmann. ISBN : 978-1-55860-737-8.
- [13] Ferraty, F., Vieu, P. (2006). *Nonparametric Functional Data Analysis : Theory and Practice*. Springer Science Business Media. ISBN : 978-0-387-38933-9.
- [14] Fréchet, M., Riesz, F. (1907). Sur la théorie des espaces abstraits. *Comptes Rendus de l'Académie des Sciences*, 144, 1414-1416.
- [15] Green, P. J., Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models : A Roughness Penalty Approach*. Chapman and Hall/CRC.
- [16] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. Springer Science Business Media. (ISBN : 978-0-387-84857-0)
- [17] Hirsch, M. W., Lacombe, J. (1997). *Differential Topology*. Springer.
- [18] Mallat, S. (2009). *A Wavelet Tour of Signal Processing : The Sparse Way (3rd Edition)*. Academic Press. ISBN :

978-0123743701.

- [19] Massart, P. (2007). Concentration inequalities and model selection : Ecole d'été de probabilités de Saint-Flour XXXIII - 2003. Lecture Notes in Mathematics, 1896. Springer-Verlag.
- [20] Papoulis, A., Pillai, S. U. (2002). Probability, Random Variables, and Stochastic Processes. McGraw-Hill. ISBN : 978-0-07-366011-0.
- [21] Ramsey, J. B., Silverman, B. W. (2005). Functional Data Analysis. Springer Science Business Media.
- [22] Ramsay, James O., Hooker, Giles et Graves, Spencer (2009), Analyse de données fonctionnelles avec R et Matlab , Springer, New York.
- [23] Reinsch, C. H. (1964). Smoothing by spline functions. Numerische Mathematik, 6(1), 177-183. DOI : 10.1007/BF01404567.
- [24] Rudin, W. (1986). Real and Complex Analysis. McGraw-Hill Education. 3rd edition. ISBN-10 : 0070542341, ISBN-13 : 978-0070542341.
- [25] Rudin, W. (1991). Functional Analysis. McGraw-Hill Science/Engineering/Math. 2nd edition. ISBN-10 : 0070542368, ISBN-13 : 978-0070542365.
- [26] Saadi, N. (2008). Estimation non paramétrique de la densité de probabilité par des fonctions orthogonales [Thèse de doctorat, Université Abderrahmane Mira de Béjaïa]. [s.l] : [s.n].
- [27] Stark, H., Woods, J. W. (1986). Probability, Random Processes, and Estimation Theory for Engineers. Prentice-Hall, Inc. ISBN 978-0-13-711706-2. OL 21138080M.
- [28] Stoica, P., Moses, R. L. (2005). Spectral Analysis of Signals. Prentice Hall. ISBN : 978-0-13-113956-3.
- [29] Szego, G. (1975). Orthogonal Polynomials. American Mathematical Society. ISBN : 978-0-8218-1933-0.
- [30] Wahba, G. (1990). Spline Models for Observational Data. SIAM. ISBN : 978-0-89871-262-8.