



**HAL**  
open science

# Techniques non-additives d'estimation de la densité de probabilité

Bilal Nehme

► **To cite this version:**

Bilal Nehme. Techniques non-additives d'estimation de la densité de probabilité. Mathématiques [math]. Université Montpellier II - Sciences et Techniques du Languedoc, 2010. Français. NNT : . tel-00576957

**HAL Id: tel-00576957**

**<https://theses.hal.science/tel-00576957>**

Submitted on 15 Mar 2011

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

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

UNIVERSITÉ MONTPELLIER II  
SCIENCES ET TECHNIQUES DU LANGUEDOC

**Thèse**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ MONTPELLIER II**

Discipline : Génie informatique, automatique et traitement du signal

École Doctorale : Information Structures Systèmes

présentée et soutenue publiquement par

**Bilal Nehme**

le 20 Décembre 2010

**Techniques non-additives d'estimation de la  
densité de probabilité.**

*Rapporteurs :* Mme. Sylvie Galichet *Professeur, Ecole Polytechnique de Savoie*  
M. Thierry Denoeux *Professeur, Université de Technologie de Compiègne*

*Examineurs :* M. Olivier Strauss *Maître de conférence, Université Montpellier II (Directeur)*  
M. William Puech *Professeur, Université Montpellier II (Président)*



# Table des matières

Introduction générale	7
<b>I Estimation précise de probabilité</b>	<b>11</b>
Introduction	11
I.1 Rappels sur la théorie de la mesure	12
I.2 Estimation non-paramétrique de densité de probabilité	14
I.2.1 Estimation naturelle de densité par les histogrammes	14
I.2.1.1 Histogramme	14
I.2.1.2 Relation entre densité de probabilité et histogramme des données	15
I.2.1.3 Estimateur de densité par histogramme	16
I.2.1.4 Propriétés statistiques de l'estimateur par histogramme	17
I.2.1.5 Utilisation de l'estimateur de densité par histogrammes	18
I.2.2 Estimateur simple de densité	21
I.2.3 Estimateur à noyau de la densité	22
I.2.3.1 Principe et définition	23
I.2.3.2 Propriétés statistiques de l'estimateur à noyau	24
I.2.3.3 Utilisation de l'estimateur de densité par noyau	27
I.3 Estimation non-paramétrique de la fonction de répartition	29
I.3.1 Estimation naturelle de la fonction de répartition par la fonction de répartition empirique	29
I.3.2 Estimation par noyau de la fonction de répartition	31
<b>II Imprécision et mesure de probabilité</b>	<b>33</b>
Introduction	33
II.1 Quelques rappels	35
II.1.1 Sous-ensembles flous	35
II.1.2 Mesures de confiances non-additives	37
II.1.2.1 Capacité	37
II.1.2.2 Intégrale de Choquet et espérance imprécise	38
II.1.2.3 Ensembles de probabilités	39
II.1.2.4 Théorie des fonctions de croyance	40
II.1.2.5 Théorie de possibilités	41
II.2 Mesure de probabilité imprécise et données imprécises	42
II.2.1 Variables aléatoires floues	43
II.2.2 Histogramme flou et distribution de probabilité floue	44
II.2.2.1 Histogrammes flous	45

II.2.2.2	Distribution de probabilité floue . . . . .	46
II.2.3	Estimation floue de densité . . . . .	48
II.3	Mesure de probabilité imprécise et données précise . . . . .	51
II.3.1	Noyaux maxitifs . . . . .	51
II.3.2	Espérance mathématique maxitive . . . . .	55
II.3.3	Estimation imprécise de la fonction de répartition . . . . .	55
II.3.3.1	Principe et définition . . . . .	56
II.3.3.2	Propriétés . . . . .	57
II.3.4	Peut-on estimer la densité de probabilité par noyau maxitif? . . . . .	58
<b>III</b>	<b>Estimation maxitive de densité de probabilité</b>	<b>59</b>
	Introduction . . . . .	59
III.1	Reformulation de l'estimateur de densité de Parzen-Rosenblatt . . . . .	60
III.1.1	Mesure empirique et estimateur de densité de Parzen-Rosenblatt . . . . .	60
III.1.2	Dérivation des distributions et estimateur de densité de Parzen-Rosenblatt . . . . .	61
III.1.3	Décomposition de la dérivée des noyaux sommatifs . . . . .	64
III.1.4	Dérivée de noyau et estimateur de densité de Parzen-Rosenblatt . . . . .	66
III.1.4.1	Estimateur de densité de Parzen-Rosenblatt . . . . .	66
III.1.4.2	Estimateur de densité de Parzen-Rosenblatt : les deux noyaux de la décomposition sont symétriques et issus de la translation d'un même noyau sommatif . . . . .	67
III.1.4.3	Estimateur de densité de Parzen-Rosenblatt : cas où le noyau est simplement symétrique . . . . .	70
III.2	Estimation maxitive de la densité . . . . .	72
III.2.1	Principe et définition de l'estimation maxitive de la densité . . . . .	72
III.2.2	Algorithme de calcul de bornes de l'estimation maxitive de la densité . . . . .	74
III.2.3	Relation entre estimation maxitive et estimation sommative de la densité . . . . .	75
III.2.4	Convergence de l'estimation maxitive de la densité . . . . .	77
III.2.4.1	Imprécision de l'intervalle d'estimation de la densité . . . . .	77
III.2.4.2	Convergence en un certain sens . . . . .	78
III.2.4.3	Comment expérimenter la convergence d'une estimation imprécise de la densité? . . . . .	80
III.3	Spécificité de l'estimation maxitive de la densité . . . . .	82
III.3.1	Première restriction sur l'intervalle d'estimation . . . . .	82
III.3.2	Deuxième restriction sur l'intervalle d'estimation . . . . .	88
<b>IV</b>	<b>Expérimentations</b>	<b>93</b>
IV.1	Décomposition sommative d'un noyau et domination de cette décomposition . . . . .	94
IV.1.1	Cas où le noyau est quelconque . . . . .	94
IV.1.2	Cas où le noyau est simplement symétrique . . . . .	101
IV.1.3	Cas où les deux noyaux de la décomposition peuvent être déduits l'un de l'autre par simple translation. . . . .	103
IV.2	Etude expérimentale de l'estimation imprécise de la densité . . . . .	104
IV.2.1	Estimation précise et imprécise de la densité : cas général . . . . .	105
IV.2.2	Estimation précise et imprécise de la densité : cas particuliers . . . . .	108

---

IV.2.3	Intégrale de l'imprécision de l'estimation imprécise de la densité . .	110
IV.2.4	Caractérisation empirique de l'estimation imprécise de la densité . .	113
IV.2.5	Inclusions de la densité convoluée dans l'estimation imprécise de la densité : cas général . . . . .	115
IV.2.6	Lien entre la variance de l'estimation précise et l'imprécision de l'estimation imprécise . . . . .	116
	<b>Conclusion générale</b>	<b>119</b>
	<b>Bibliographie</b>	<b>123</b>



# Table des figures

I.1	Histogramme de 100 observations issues d'une loi normale $\mathcal{N}(0, 1)$ construit sur une partition de granularité $\Delta = 0.8$ d'un intervalle de référence $\Omega = [-5, 5]$ . . . . .	15
I.2	Histogramme de 100000 observations construit avec une granularité $\Delta = 0.2$ . . . . .	16
I.3	Histogramme de 300 observations construit avec une granularité $\Delta = 2.5$ (a) et $\Delta = 0.1$ (b). . . . .	19
I.4	Histogramme de 300 observations construit sur la portion $\Omega = [-5, 5]$ (a) et $\Omega = [-4, 6]$ (b). . . . .	20
I.5	Superposition de quatre noyaux sommatifs. . . . .	23
I.6	L'estimateur de densité de Parzen-Rosenblatt construit à partir de 1000 observations en utilisant le noyau cosinus avec une largeur de bande $\Delta = 0.2$ (a) et $\Delta = 5$ (b). . . . .	28
I.7	Variation sur la forme du noyau $\kappa$ pour la méthode d'estimation par noyau. . . . .	29
II.1	Un noyau sommatif $\kappa$ vérifie la propriété de sommativité (a) et un noyau maxitif $\pi$ vérifie la propriété de maxitivité (b). . . . .	52
III.1	La dérivée d'un noyau sommatif $\kappa$ comme la combinaison linéaire de deux noyaux sommatifs $\eta^+$ et $\eta^-$ . . . . .	65
III.2	Décomposition sommative de la dérivée première de $\kappa$ : cas où les deux noyaux $\eta^+$ et $\eta^-$ sont symétriques et issus de la translation d'un même noyau sommatif. . . . .	69
III.3	Décomposition sommative de la dérivée première de $\kappa$ : cas où le noyau sommatif $\kappa$ est simplement symétrique. . . . .	71
IV.1	Les deux noyaux sommatifs $\eta^-$ (a) et $\eta^+$ (b) intervenant dans la décomposition sommative du noyau d'Epanechnikov. . . . .	95
IV.2	Le noyau sommatif $\eta^+(x) = 2x\mathbb{1}_{[0,1]}(x)$ et l'intervalle de confiance $I_{u=0.5}$ . . . . .	96
IV.3	Les deux noyaux maxitifs $\pi^-$ (a) et $\pi^+$ (b) intervenant dans la domination objective de la décomposition sommative du noyau d'Epanechnikov. . . . .	97
IV.4	Les deux noyaux maxitifs $\pi^-$ (a) et $\pi^+$ (b) intervenant dans la domination subjective de la décomposition sommative du noyau d'Epanechnikov. . . . .	98
IV.5	Les deux noyaux sommatifs $\eta^-$ (a) et $\eta^+$ (b) intervenant dans la décomposition sommative du noyau cosinus symétrisé. . . . .	99
IV.6	Le noyau sommatif $\eta^+(x) = \frac{\pi}{2} \sin(x\pi)\mathbb{1}_{[0,1]}(x)$ et l'intervalle de confiance $I_{u=0.25}$ (rsp. $I_{u=0.75}$ ) (a) (rsp. (b)). . . . .	100



IV.7	Les deux noyaux maxitifs $\pi^-$ (a) et $\pi^+$ (b) intervenant dans la domination objective de la décomposition sommative du noyau cosinus symétrisé. . . . .	101
IV.8	Le noyau sommatif de translation $\eta$ (a) intervenant dans la décomposition sommative du noyau d'Epanechnikov et le noyau maxitif $\pi$ (b) dominant, via une domination objective, le noyau $\eta$ . . . . .	102
IV.9	Le noyau sommatif de translation $\eta$ (a) intervenant dans la décomposition sommative du noyau cosinus symétrisé et le noyau maxitif $\pi$ (b) dominant, via une domination objective, le noyau $\eta$ . . . . .	103
IV.10	Superposition de l'estimation précise (noyau cosinus symétrisé) et l'estimation imprécise (domination plus spécifique). . . . .	105
IV.11	Superposition de l'estimation imprécise et l'estimation précise pour $\Delta = 5$ (a) et $\Delta = 0.05$ (b). Le code de couleurs utilisé est le même que dans la Figure. IV.10. . . . .	106
IV.12	Superposition de l'estimation précise (noyau d'Epanechnikov) et l'estimation imprécise de la densité (domination la plus spécifique). . . . .	107
IV.13	Superposition de l'estimation précise (noyau d'Epanechnikov) et l'estimation imprécise de la densité (domination moins spécifiques). . . . .	108
IV.14	Première restriction : superposition de l'estimation précise (noyau cosinus symétrisé) et l'estimation imprécise (domination plus spécifique). . . . .	109
IV.15	Deuxième restriction : superposition de l'estimation précise (noyau cosinus symétrisé) et l'estimation imprécise (domination plus spécifique). . . . .	110
IV.16	Valeur théorique et valeur calculée de l'intégrale de l'imprécision en fonction : (a) de la largeur de bande et (b) du nombre d'observations. . . . .	111
IV.17	Valeur théorique et valeur calculée de l'intégrale médiane, inférieure et supérieure de l'imprécision de l'estimation imprécise en fonction : (a) de la largeur de bande et (b) du nombre d'observations. . . . .	112
IV.18	Superposition des 100 estimations précises (noyau d'Epanechnikov), estimations imprécises (domination moins spécifique) et densités convoluées. . . . .	114
IV.19	Courbes de performance de l'estimation imprécise en fonction de la largeur de bande. . . . .	115
IV.20	Superposition de quatre distances $d_H, d_M, d_S$ et $d_1$ et la valeur constante $\beta$ en fonction du nombre d'observations. . . . .	116
IV.21	Le nuage de valeurs de $\gamma(n, \Delta)\mathbb{E}(\Gamma_{(\kappa_\Delta)}^n)$ en fonction de $var(\hat{f}_{\kappa_\Delta}^n)$ . . . . .	118

# Introduction générale

L'estimation de la densité de probabilité sous-jacente à un ensemble fini d'observations est un problème fondamental qui a fait l'objet d'une très vaste littérature [106, 115, 104, 10]. On retrouve cette problématique dans de nombreux domaines des sciences et techniques tels que le traitement du signal et des images, la mécanique, la robotique, le contrôle automatique, ... et toutes les techniques utilisant des mesures du réel. Le principe même de mesure est fortement associé à celui des variations aléatoires. Ces variations sont liées soit aux phénomènes que l'on souhaite mesurer (dans le cas de mesures de radioactivité dans les images médicales) soit à des phénomènes physiques utilisés pour mesurer (dans le cas des mesures inertielles par exemple) ou transmettre les signaux associés aux phénomènes étudiés. La modélisation de ces variations aléatoires est traditionnellement réalisée par une distribution de probabilité. Dans de très rares cas, une analyse fine du processus permet de définir cette distribution. Mais généralement on doit l'identifier à partir d'un nombre fini d'observations, identification qui consiste généralement à estimer la densité de probabilité en tout point de l'espace considéré. Nous considérons ici des mesures réelles et restreignons notre discours à un sous-ensemble des réels.

On trouve dans la littérature deux types d'approches d'estimations de la densité de probabilité : l'approche paramétrique et l'approche non-paramétrique. L'approche paramétrique a comme inconvénient principal de nécessiter une connaissance préalable sur la loi de probabilité du phénomène aléatoire que l'on étudie. L'approche non-paramétrique estime la densité de probabilité directement à partir de l'information disponible sur l'ensemble d'observations. On dit souvent que dans cette approche *les données parlent d'elles mêmes*. Nous nous intéressons ici plutôt aux approches non-paramétriques.

Dans toute approche d'estimation ou d'identification, la fiabilité du résultat produit est fortement liée au nombre d'observations permettant de réaliser l'identification. La garantie d'une estimation fiable et sans biais n'est obtenue que pour un nombre très grand (en théorie infini) d'observations. Lorsque le nombre d'observations est insuffisant, l'estimation est incertaine sans que cette incertitude soit vraiment mesurable. Si la qualité d'une estimation paramétrique est fortement liée à la validité de l'hypothèse faite sur la loi de probabilité identifiée, celle de l'approche non-paramétrique dépend elle du choix d'une fonction, appelée noyau, définissant un voisinage pondéré autour de tout point du domaine considéré. Une variation de choix entraîne une variation de l'estimation, et ce particulièrement lorsque le nombre d'observations est petit. Lorsque le nombre d'observations est grand, le choix du noyau a une incidence moindre sur l'identification.

En situation non-asymptotique (faible nombre de données), le côté arbitraire du choix du noyau peut avoir une incidence importante sur le processus dans lequel cette modélisation est utilisée. Pour donner un exemple commun, supposons que la distribution de probabilité serve dans un processus d'aide à la décision en imagerie médicale. Si deux

choix de voisinage conduisent à deux décisions médicales différentes (voire opposées), il semble nécessaire de donner cette information au praticien. Une façon de répondre à cette nécessité serait d'utiliser un ensemble de fonctions de voisinages compatibles avec les données à analyser et d'estimer l'ensemble des distributions de probabilités associées à cet ensemble de fonctions de voisinage. Cet ensemble de fonctions de voisinage semble devoir être convexe par nature. En effet, si deux fonctions  $\phi$  et  $\psi$  sont envisageables en tant que noyaux de voisinage, alors la fonction  $\alpha\phi + (1 - \alpha)\psi$  ( $\alpha \in [0, 1]$ ) devrait l'être aussi. Pour représenter de façon synthétique cet ensemble de densité de probabilités, on peut utiliser une densité de probabilité imprécise au sens de De Campos et al. [18], c'est-à-dire un ensemble de deux fonctions  $(\underline{f}, \overline{f})$  qui sont les bornes inférieures et supérieures de l'ensemble des densités à estimer (mais qui ne sont elles-mêmes pas des densités). L'estimation de ces deux fonctions bornantes peut faire l'objet d'une procédure de type Monte-Carlo . . . mais ce type de méthode se révèle extrêmement coûteux en temps de calcul. De plus les bornes de l'intervalle sont estimées et non pas calculées.

Une autre façon de procéder a été proposée par Loquin et Strauss dans [72]. Elle consiste à représenter un ensemble de noyaux de voisinage (appelés *noyaux sommatifs*) par un nouveau type de voisinage appelé *noyau maxitif*. Un noyau maxitif peut être vu comme représentant un ensemble convexe de noyaux sommatifs. Cette représentation s'appuie fortement sur la théorie des probabilités imprécises. En étendant aux noyaux maxitifs les opérations d'estimations réalisées habituellement avec des noyaux sommatifs, Loquin et Strauss ont proposé de définir un estimateur imprécis de la fonction de répartition à la *Parzen-Rosenblatt*. Cette extension utilise une généralisation de l'opérateur espérance mathématique connue sous le nom d'intégrale de Choquet. Elle permet de calculer simplement les deux fonctions de répartition qui sont les bornes inférieures et supérieures de l'ensemble des fonctions de répartition qui auraient été estimées en utilisant l'ensemble des noyaux sommatifs représenté par le noyau maxitif considéré. Cette approche présente l'intérêt notoire, par rapport à l'approche basée sur Monte-Carlo, d'avoir une complexité comparable à celle de l'estimation à noyau classique et de permettre de calculer de façon exacte les bornes de l'ensemble des fonctions de répartitions au lieu d'en réaliser une estimation. Il pourrait donc sembler tout à fait pertinent d'utiliser une approche analogue pour calculer les deux fonctions  $(\underline{f}, \overline{f})$  délimitant l'ensemble des densités de probabilités estimées à partir d'un ensemble convexe de noyaux. Cependant, pour une raison technique que nous allons expliquer dans ce manuscrit, il n'est pas possible d'utiliser directement cette approche.

Dans ce travail de thèse, nous proposons de contourner cette difficulté en utilisant la théorie des distributions de Schwartz [98] et le lien particulier unissant fonction de densité et fonction de répartition. L'estimateur que nous obtenons est, comme l'estimateur de fonction de répartition de Loquin et Strauss, un estimateur imprécis, ou plus exactement un estimateur des bornes de l'ensemble des estimations qui auraient été obtenues à partir d'un ensemble convexe de noyaux sommatifs. Nous donnons les fondements de cette approche. Nous présentons un certain nombre de propriétés intéressantes de cet estimateur ainsi que les limites de cette approche. Nous proposons aussi des algorithmes pratiques de faible complexité pour réaliser les estimations proposées.

Le présent mémoire est subdivisé en quatre chapitres. Le chapitre I fait un état de l'art des techniques d'estimation des fonctions de densité et de répartition de la probabilité sous-jacente à un ensemble d'observations. Nous présentons les techniques paramétriques et non-paramétriques en insistant particulièrement sur ces dernières qui font l'objet de

ce travail de thèse. Nous présentons les propriétés statistiques de chaque méthode et particulièrement ce qui concerne la convergence de ces estimateurs. Notre approche produisant une estimation imprécise de la densité, nous proposons, dans le chapitre II, un état de l'art des relations entre imprécision et estimation de mesures de probabilités. Après quelques rappels sur ce qu'il est convenu d'appeler imprécision, nous présentons plusieurs techniques aboutissant à l'estimation d'une mesure de probabilité imprécise à partir d'un ensemble fini d'observations. Notons que hormis la technique proposée par Loquin et Strauss précédemment citée, la plupart de ces techniques propagent, sur l'estimation de la probabilité, l'imprécision des observations et non pas l'imprécision du modèle d'estimation. Le chapitre III présente le cœur de ce travail de thèse, c'est-à-dire la construction d'un estimateur imprécis de la densité de probabilité sous-jacente à un ensemble d'observations précises. L'imprécision dont nous parlons ici est celle que nous avons déjà évoqué, c'est-à-dire celle induite par la méconnaissance du noyau permettant d'estimer au mieux la densité de probabilité (qui peut finalement être vue comme un paramètre de cette méthode non-paramétrique). Cette construction s'appuie sur une décomposition particulière de la dérivée d'un noyau sommatif. Du à la particularité de cette construction, nous proposons plusieurs techniques d'estimation utilisant certaines propriétés des noyaux comme leur symétrie par exemple. Pour chaque méthode d'estimation, nous donnons un algorithme de faible complexité permettant de la programmer. Nous donnons aussi un certain nombre de propriétés de cet estimateur ainsi que des mesures de distance permettant de tester la qualité des estimations obtenues par rapport à une vérité terrain connue. Dans le chapitre IV, nous proposons un certain nombre d'expérimentations des méthodes que nous proposons, certaines pour illustrer des propriétés qui ont été démontrées au chapitre précédent, d'autres pour appréhender certaines conjectures comme l'aptitude de cet estimateur imprécis à quantifier sa propre erreur d'estimation ou encore mesurer la convergence de cette estimation.



# Chapitre I

## Estimation précise de probabilité

### Introduction

Pour caractériser un phénomène aléatoire sous-jacent à un ensemble fini d'observations, il est souvent pertinent d'estimer la *fonction de répartition* (probabilité cumulée) ou la *fonction de densité* (densité de probabilité). Ces deux fonctions sont reliées par le fait que la fonction de densité est la dérivée de la fonction de répartition.

Même si les fonctions de répartition et de densité caractérisent toutes les deux la loi de probabilité d'une variable aléatoire, la densité a un net avantage sur le plan visuel. En fait, la forme d'une densité est beaucoup plus facile à interpréter que celle de la fonction de répartition. Par exemple, voir la symétrie ou la multimodalité de la loi d'une variable aléatoire est beaucoup plus facile sur une fonction de densité que sur une fonction de répartition. Une estimation de densité peut aussi aider dans le choix d'un modèle paramétrique approprié pour la loi d'une variable aléatoire.

Pour estimer ces deux fonctions de densité et de répartition, on ne dispose généralement que d'un ensemble fini d'observations issu d'une même variable aléatoire. Le problème statistique fondamental consiste à utiliser cet ensemble d'observations pour construire un estimateur de la fonction de densité et la fonction de répartition. Il existe deux approches très complémentaires pour réaliser ces estimations : l'approche d'*estimation paramétrique* et l'approche d'*estimation non-paramétrique*. L'approche d'estimation paramétrique suppose que les données sont issues d'une loi de probabilité de forme connue dont seuls les paramètres sont inconnus. Dans ce cas, l'estimation de densité se résume à l'estimation des paramètres de la distribution. Si la loi de probabilité est inconnue, ou s'il s'agit justement de trouver la forme de cette loi sans à priori pour ensuite en réaliser une estimation paramétrique, on doit se tourner vers une méthode non-paramétrique dans laquelle les données *parlent d'elles mêmes*.

Dans ce chapitre, nous nous plaçons dans le cadre de l'estimation non-paramétrique. En section I.1, nous introduisons quelques rappels sur la théorie de la mesure. Nous nous concentrons, en section I.2, sur trois méthodes d'estimation de la densité : la méthode d'*estimation par histogramme* (estimation naturelle de la densité), la méthode d'*estimation par histogramme mobile* (estimateur simple de la densité) et la méthode d'*estimation par noyau* (*estimateur de densité Parzen-Rosenblatt*) qui peut être vue comme une extension de la méthode d'estimation par histogramme. Nous présentons également, dans cette section, les propriétés statistiques de chaque méthode d'estimation. En section I.3, nous présentons deux méthodes d'estimation non-paramétrique de la fonction

de répartition : la *fonction de répartition empirique* (estimation naturelle de la fonction de répartition) et la méthode d'*estimation par noyau*. Nous abordons aussi les propriétés statistiques de chaque méthode d'estimation.

## I.1 Rappels sur la théorie de la mesure

Nous introduisons, dans cette section, quelques outils et définitions utilisés dans la théorie de la mesure. Nous considérons, tout au long de ce manuscrit, que les observations appartiennent à un sous-ensemble convexe et compact<sup>1</sup>  $\Omega$  de  $\mathbb{R}$ . Nous appelons  $\Omega$  l'*ensemble de référence*.  $\mathcal{P}(\Omega)$  est l'ensemble de partie de  $\Omega$ .

**Définition I.1 (Topologie ([4] page 2))** Une topologie sur  $\Omega$  est donnée par une famille de parties de  $\Omega$ , notée  $\mathcal{E}$ , contenant  $\emptyset$  et  $\Omega$ , stable par union quelconque ( $\forall A, B \in \mathcal{E}, A \cup B \in \mathcal{E}$ ) et stable par intersection finie ( $\forall A, B \in \mathcal{E}, A \cap B \in \mathcal{E}$ ). Le couple  $(\Omega, \mathcal{E})$  est appelé *espace topologique*.

**Définition I.2 (Tribu ([4] page 2))** Soit  $\mathcal{A}$  une famille de parties de  $\Omega$  ( $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ ), la famille  $\mathcal{A}$  est une *tribu* (ou une  $\sigma$ -algèbre) sur  $\Omega$  si  $\mathcal{A}$  contient  $\emptyset$  et  $\Omega$ , est stable par union dénombrable ( $\forall \{A_i\}_{i \in \mathbb{N}} \subset \mathcal{A}, \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$ ), stable par intersection dénombrable ( $\forall \{A_i\}_{i \in \mathbb{N}} \subset \mathcal{A}, \bigcap_{i \in \mathbb{N}} A_i \in \mathcal{A}$ ) et stable par passage au complémentaire ( $\forall A \in \mathcal{A}, A^c \in \mathcal{A}$ , où  $A^c$  est le complémentaire de  $A$  de  $\Omega$ ). Le couple  $(\Omega, \mathcal{A})$  est appelé *espace mesurable*.

La tribu la plus utilisée en statistique est la  $\sigma$ -algèbre borélienne (la plus petite  $\sigma$ -algèbre qui contient les ensembles ouverts d'un espace topologique). Nous appelons  $\mathcal{B}(\Omega)$  la tribu borélienne sur l'espace topologique  $(\Omega, \mathcal{E})$ .

Soit  $(\Omega, \mathcal{A})$  un espace mesurable, une application  $m$  de  $\mathcal{A}$  à valeurs dans  $\mathbb{R} \cup (+\infty)$  est dite  $\sigma$ -additive si pour toute famille  $\{A_i\}_{i \in \mathbb{N}}$  d'éléments de  $\mathcal{A}$

$$m\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} m(A_i).$$

**Définition I.3 (Mesure, probabilité ([4] page 13))** Une mesure est une application  $m$  de  $\mathcal{A}$  à valeurs dans  $\mathbb{R}^+ \cup (+\infty)$ ,  $\sigma$ -additive telle que  $m(\emptyset) = 0$ . Si une mesure  $m$  sur  $\mathcal{A}$  est telle que  $m(\mathcal{A}) < \infty$ , alors  $m$  est une *mesure finie*.

On appelle *probabilité* une mesure finie  $P$  sur  $\mathcal{A}$  telle que  $P(\mathcal{A}) = 1$ . En terme probabiliste, le couple  $(\Omega, \mathcal{A})$  est appelé *espace probabilisable*. Le triplet  $(\Omega, \mathcal{A}, m)$  (resp.  $(\Omega, \mathcal{A}, P)$ ) est appelé *espace mesuré* (resp. *probabilisé*).

Soient  $(\Phi, \mathcal{R})$  et  $(\Upsilon, \mathcal{S})$  deux espaces mesurables et  $G$  une application de  $\Phi$  dans  $\Upsilon$ ,  $G$  est une *application mesurable* si

$$\forall A \in \mathcal{S}, G^{-1}(A) \in \mathcal{R},$$

où  $G^{-1}(A)$  est l'image réciproque de  $A$  par  $G$  ([4] page 5).

---

1. Un ensemble compact est un ensemble fermé et borné ([5] page 308).

**Définition I.4 (Ensembles aléatoires ([78] page 2))** Soit  $\mathcal{G}(\Omega)$  la famille des sous-ensembles fermés de  $\Omega$  et soit  $\mathcal{S}$  la tribu engendrée par les sous-ensembles de  $\mathcal{G}$  tels que leur intersection avec un compact  $K$  est non-vide, c'est-à-dire  $\mathcal{S} = \{E \in \mathcal{G} : K \cap E \neq \emptyset\}$ . Soient  $(\Omega, \mathcal{R}, P)$  un espace probabilisé et  $(\mathcal{G}, \mathcal{S})$  un espace mesurable, un ensemble aléatoire est une application mesurable de  $(\Omega, \mathcal{R})$  dans  $(\mathcal{G}, \mathcal{S})$ .

Soit  $(\Omega, \mathcal{A})$  un espace probabilisable et soit  $\mathcal{B}(\Omega)$  la tribu borélienne de  $\Omega$ , une variable aléatoire est une application  $X$  définie de  $\Omega$  à valeurs dans  $\mathbb{R}$  qui est mesurable, c'est-à-dire  $\forall A \in \mathcal{B}(\Omega), X^{-1}(A) \in \mathcal{A}$ .

La fonction de répartition de la variable aléatoire  $X$ , notée  $F_X$ , est définie par :

$$\forall x \in \Omega, F_X(x) = P(X \leq x),$$

la valeur prise de la fonction de répartition au point  $x$  est la probabilité de l'événement  $] -\infty, x]$ .

Dans la suite de ce manuscrit, nous appelons  $L(\Omega)$  l'ensemble des fonctions définies sur  $\Omega$  à valeurs dans  $\mathbb{R}$ ,  $L^1(\Omega)$  l'ensemble des fonctions mesurables intégrables au sens de Lebesgue sur  $\Omega$  et  $L^1_{loc}(\Omega)$  l'ensemble des fonctions de  $L^1(\Omega)$  qui sont intégrables sur tout compact  $K$  de  $\Omega$ .

**Définition I.5 (Densité de probabilité)** Soit  $X$  une variable aléatoire et  $F_X$  sa fonction de répartition, s'il existe une fonction  $f_X$  positive de  $L^1(\Omega)$  telle que

$$\forall x \in \Omega, F_X(x) = \int_{-\infty}^x f_X(u) du,$$

alors  $f_X$  s'appelle la densité de probabilité de la variable aléatoire  $X$ .  $f_X$  vérifie :

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

Lorsqu'on connaît la densité de probabilité  $f_X$  de  $X$ , il est possible de calculer la probabilité d'appartenance d'une variable aléatoire  $X$  à n'importe quel ensemble  $A \subseteq \Omega$  :

$$P(X \in A) = \int_A f_X(x) dx.$$

**Définition I.6 (Espérance mathématique ([4] page 53))** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et soit  $X$  une variable aléatoire de  $(\Omega, \mathcal{A}, P)$  dans  $(\Omega, \mathcal{B}(\Omega))$ . Si  $X \in L^1(\Omega)$ , l'espérance mathématique de la variable aléatoire  $X$ , notée  $\mathbb{E}(X)$ , est définie par :

$$\mathbb{E}(X) = \int_{\Omega} X dP.$$

Une propriété intéressante découle de cette définition :

**Propriété I.7 (Espérance mathématique d'une fonction d'une variable aléatoire)**

Soit  $X$  une variable aléatoire et soit  $s \in L^1(\Omega)$  une fonction de  $X$ , l'espérance mathématique de la variable aléatoire  $Y = s(X)$  est donnée par :

$$\mathbb{E}(Y) = \mathbb{E}(s(X)) = \int_{\Omega} s(x) f_X(x) dx.$$

Dans la suite de ce manuscrit, la variable aléatoire sera toujours  $X$ . Nous noterons donc  $F$  (resp.  $f$ ) et non  $F_X$  (resp.  $f_X$ ) la fonction de répartition (resp. densité) de cette variable aléatoire.



## I.2 Estimation non-paramétrique de densité de probabilité

L'avantage principal de l'estimation non-paramétrique de la densité de probabilité sous-jacente à un ensemble fini d'observations est de ne pas nécessiter d'hypothèses à priori sur l'appartenance de cette densité à une famille de lois connues. L'estimation ne concerne pas les paramètres permettant de sélectionner une loi, mais directement la fonction elle-même (d'où le terme non-paramétrique). Il existe une littérature abondante sur l'estimation non-paramétrique de la densité. Une bonne introduction à cette problématique peut être trouvée dans les livres de Bosq et Lecoute [10], Silverman [104] et Tsybakov [115].

Dans la suite de ce manuscrit, l'ensemble de référence  $\Omega$  est un intervalle et  $e_{min}$  (resp.  $e_{max}$ ) est le plus petit (resp. plus grand) élément de  $\Omega$ .

### I.2.1 Estimation naturelle de densité par les histogrammes

L'histogramme est le plus ancien des estimateurs non-paramétriques de densité. D'après ([68] page 182), l'origine des histogrammes est attribuée à John Graunt au XVII<sup>e</sup> siècle répondant à l'objectif d'une représentation de la distribution de données. À ce titre, il peut être considéré comme un estimateur de la densité de probabilité sous-jacente à un ensemble fini d'observations. Nous voyons, dans cette section, comment obtenir cette dernière propriété. Nous discutons ensuite des propriétés statistiques de l'estimateur de densité par histogramme ainsi que des problèmes engendrés par son utilisation pour représenter la densité de probabilité.

#### I.2.1.1 Histogramme

Construire un histogramme à partir d'un ensemble d'observations (précises)  $(x_1, \dots, x_n)$  consiste à partitionner l'intervalle de référence  $\Omega = [e_{min}, e_{max}]$  en  $p \in \mathbb{N}$  cellule  $A_k$ ,  $k \in \{1, \dots, p\}$ , et à compter le nombre d'observations appartenant à chaque cellule  $A_k$ . Si toutes les cellules de l'histogramme ont la même largeur, on dit que l'histogramme est régulier. On note  $\Delta \in \mathbb{R}^+$ , la largeur des cellules qui est alors appelée le *pas* ou la *granularité* de l'histogramme. La valeur de  $\Delta$  est donnée par :

$$\Delta = \frac{e_{max} - e_{min}}{p}. \quad (\text{I.1})$$

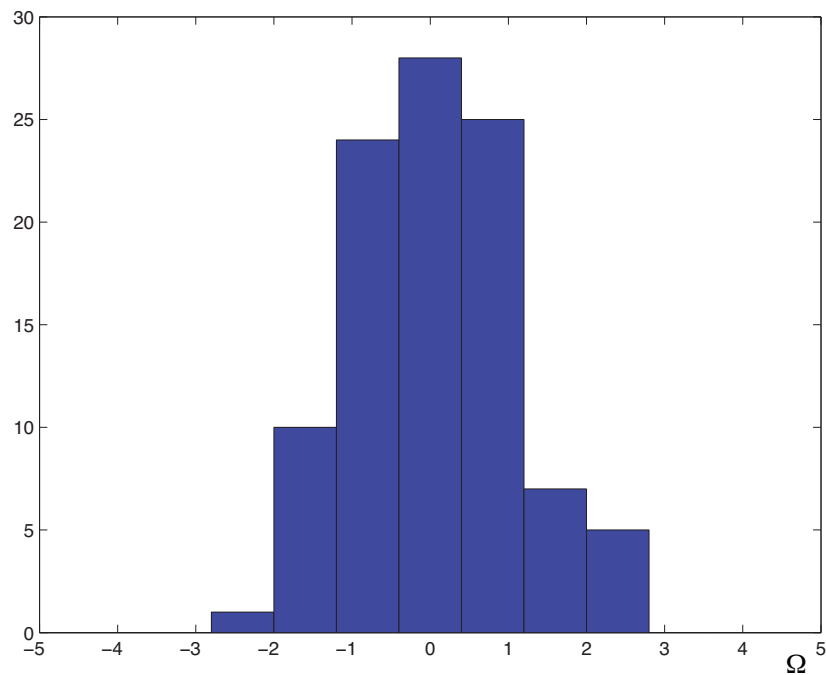
Le nombre d'observations appartenant à chaque cellule  $A_k$  est appelé *accumulateur* de la cellule  $A_k$  est noté  $\text{Acc}_k$ . Il est défini ([69] page 46) par :

$$\text{Acc}_k = \sum_{i=1}^n \mathbb{1}_{A_k}(x_i), \quad (\text{I.2})$$

où  $\mathbb{1}_{A_k}(x)$  est la fonction caractéristique de l'ensemble  $A_k$  définie par :

$$\mathbb{1}_{A_k}(x) = \begin{cases} 1 & \text{si } x \in A_k, \\ 0 & \text{sinon.} \end{cases} \quad (\text{I.3})$$

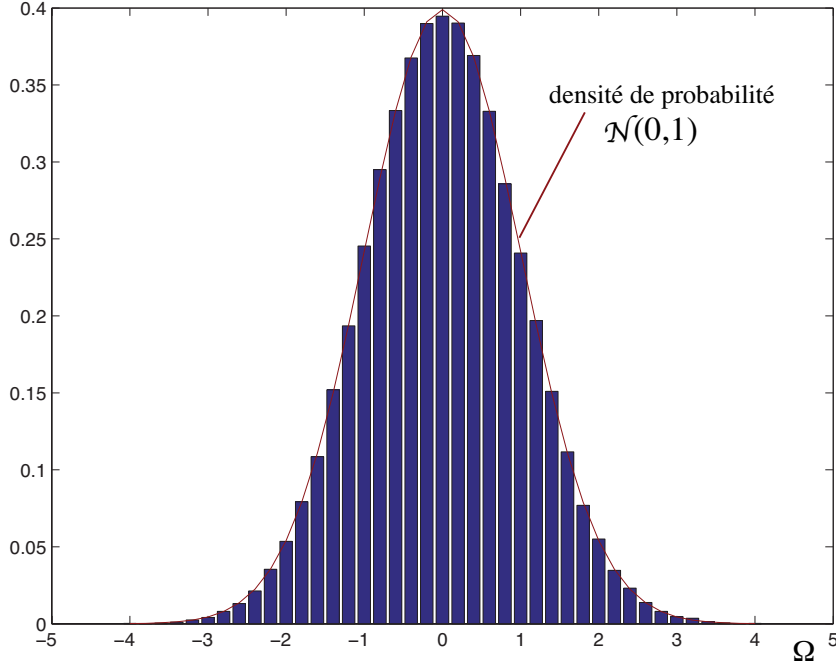
La Figure I.1 présente un histogramme de 100 observations tirées aléatoirement d'une loi normale centrée réduite  $\mathcal{N}(0, 1)$ . Ces observations sont réparties sur un intervalle de référence  $\Omega = [-5, 5]$ . La granularité de l'histogramme est fixée à  $\Delta = 0.8$ .



**Figure I.1 :** Histogramme de 100 observations issues d'une loi normale  $\mathcal{N}(0, 1)$  construit sur une partition de granularité  $\Delta = 0.8$  d'un intervalle de référence  $\Omega = [-5, 5]$ .

### I.2.1.2 Relation entre densité de probabilité et histogramme des données

Histogramme et densité de probabilité sont liés par des conditions aux limites : *une densité de probabilité peut être vue comme la limite d'un histogramme lorsque le nombre d'observations tend vers l'infini et que la granularité de l'histogramme tend vers zéro*. La Figure I.2 illustre, cette relation, en considérant le même processus d'observations qu'en section I.2.1.1 avec un grand nombre d'observations ( $n = 100000$ ) et une plus faible granularité ( $\Delta = 0.2$ ). En comparant la Figure I.1 avec la Figure I.2, on voit clairement que l'allure de cet histogramme se rapproche de l'allure de la densité de probabilité ( $\mathcal{N}(0, 1)$ ) quand le nombre d'observations  $n$  augmente et la granularité  $\Delta$  diminue.



**Figure I.2** : Histogramme de 100000 observations construit avec une granularité  $\Delta = 0.2$ .

### I.2.1.3 Estimateur de densité par histogramme

Soit un ensemble fini d'observations  $(x_1, \dots, x_n)$  de  $n$  variables aléatoires  $(X_1, \dots, X_n)$  indépendantes (au sens de probabilité) et identiquement distribuées (de même loi de probabilité) (i.i.d) de densité de probabilité commune  $f$ . Pour tout  $k \in \{1, \dots, p\}$ , soit  $\text{Acc}_k$  l'accumulateur associé à la cellule  $A_k$ , défini par l'expression (I.2), la probabilité de  $A_k$  (basée sur les observations  $x_i$ ), notée  $P(A_k)$ , est donnée par :

$$P(A_k) = \frac{\text{Acc}_k}{n}. \quad (\text{I.4})$$

En émettant l'hypothèse, généralement irréaliste, que les observations se répartissent uniformément dans la cellule  $A_k$  (de largeur  $\Delta$ ), on peut alors construire un estimateur de la densité  $f$ , pour tout  $x \in \Omega$ , par :

$$\hat{f}_\Delta^n(x) = \frac{1}{\Delta} \sum_{k=1}^p P(A_k) \mathbb{1}_{A_k}(x). \quad (\text{I.5})$$

Cet estimateur peut aussi s'écrire

$$\hat{f}_\Delta^n(x) = \frac{1}{n\Delta} \sum_{k=1}^p \text{Acc}_k \mathbb{1}_{A_k}(x). \quad (\text{I.6})$$

Dans la suite, nous émettons l'hypothèse que les cellules  $A_k$ ,  $\forall k \in \{1, \dots, p\}$  forment une partition de  $\Omega$  (c'est-à-dire  $\Omega = \cup_{k=1}^p A_k$  et  $\forall i, j \in \mathbb{N} : i \neq j, A_i \cap A_j = \emptyset$ ) et définissons pour chaque cellule  $A_k$ , son *centre*  $a_k$  telles que  $\forall k \in \{1, \dots, p\}$ ,  $A_k = [a_k - \frac{\Delta}{2}, a_k + \frac{\Delta}{2}]$  et  $\forall k \in \{1, \dots, p-1\}$ ,  $a_{k+1} = a_k + \Delta$ .

### I.2.1.4 Propriétés statistiques de l'estimateur par histogramme

Nous présentons, dans cette section, les propriétés statistiques de l'estimateur par histogramme  $\hat{f}_\Delta^n$  défini par l'expression (I.6) (pour une étude plus détaillée voir, par exemple, les livres de Bosq et Lecoutre [10] et de Simonoff [106]).

En statistiques, il est nécessaire de mesurer la qualité d'un estimateur. Pour cela, on évalue, d'une part, l'écart entre la moyenne de l'estimateur et la densité à estimer, ce critère d'évaluation est appelé *biais*<sup>2</sup>, et d'autre part, la *variance* de l'estimateur (due au caractère aléatoire d'observations) qui caractérise la dispersion des valeurs de l'estimateur dans l'ensemble d'observations. On essaye généralement de réduire au mieux ces deux quantités.

Le biais de l'estimateur  $\hat{f}_\Delta^n$  (I.6) est donné ([106] page 16), pour tout  $x \in (a_k, a_{k+1}]$ ,  $\forall k \in \{1, \dots, p\}$ , par :

$$\begin{aligned} \text{biais}(\hat{f}_\Delta^n(x)) &= \mathbb{E}(\hat{f}_\Delta^n(x)) - f(x) \\ &= \frac{1}{2}f'(\Delta - 2(x - a_k)) + \mathcal{O}(\Delta^2), \end{aligned} \quad (\text{I.7})$$

où  $\mathcal{O}$  est un terme résiduel et  $f'$  est la dérivée de  $f$ .  $f'$  doit être une fonction de  $L(\Omega)$  absolument continue<sup>3</sup> et carrée intégrable<sup>4</sup>.

La variance de l'estimateur est donnée ([106] page 16), pour tout  $x \in \Omega$ , par :

$$\begin{aligned} \text{var}(\hat{f}_\Delta^n(x)) &= \mathbb{E}\left(\left(\hat{f}_\Delta^n(x)\right)^2\right) - \left(\mathbb{E}(\hat{f}_\Delta^n(x))\right)^2 \\ &= \frac{f(x)}{n\Delta} + \mathcal{O}(n^{-1}). \end{aligned} \quad (\text{I.8})$$

Cette variance tend vers zéro quand le produit  $n\Delta$  tend vers l'infini quand le nombre d'observation  $n$  tend vers l'infini.

Afin d'apprécier la qualité de l'estimateur, il est usuel d'évaluer la distance entre l'estimateur et la densité à estimer. La distance la plus couramment utilisée est celle définie par la moyenne du carré de la valeur absolue de leur différence. Elle porte le nom d'*erreur quadratique moyenne* (MSE)<sup>5</sup>.

La convergence en *moyenne quadratique* de l'estimateur  $\hat{f}_\Delta^n$  (I.6) a été établie par Lecoutre [67]. Nous avons, d'après ([106] page 16), pour tout  $x \in (a_k, a_{k+1}]$  :

$$\begin{aligned} \text{MSE}(\hat{f}_\Delta^n(x)) &= \text{var}(\hat{f}_\Delta^n(x)) + \text{biais}^2(\hat{f}_\Delta^n(x)) \\ &= \frac{f(x)}{n\Delta} + \frac{f'(x)^2}{4}(\Delta - 2(x - a_k))^2 + \mathcal{O}(n^{-1}) + \mathcal{O}(\Delta^3). \end{aligned} \quad (\text{I.9})$$

Cette erreur quadratique moyenne tend vers zéro quand  $\Delta$  tend vers zéro et  $n\Delta$  tend vers l'infini quand  $n$  tend vers l'infini. Ce critère d'erreur quadratique moyenne est un critère

2. Le biais d'un estimateur  $\hat{s}^n$  de  $s \in L(\Omega)$  est donné par :  $\text{biais}(\hat{s}^n) = \mathbb{E}(\hat{s}^n) - s$  ([68] page 99).

3. Soit  $A = [a, b]$  un intervalle de  $\Omega$  et soit  $s$  une fonction de  $L(\Omega)$ . On dit que la fonction  $s$  est absolument continue sur  $A$  si, pour tout  $\epsilon > 0$ , il existe un  $\iota > 0$  tel que, pour toute suite  $(a_n, b_n)_{n \in \mathbb{N}}$  de sous-intervalle de  $A$  d'intérieurs disjoints :  $\sum_{n \geq 0} (b_n - a_n) < \iota \Rightarrow \sum_{n \geq 0} (|s(b_n) - s(a_n)|) < \epsilon$  ([84] page 12).

4. Une fonction  $s \in L^1(\Omega)$  est dite carrée intégrable si :  $\int_{-\infty}^{\infty} |s(u)|^2 du < \infty$  ([62] page 79).

5. En anglais Mean Squared Error.

local. On lui préfère généralement un critère plus global obtenu en l'intégrant sur tout le domaine  $\Omega$ . Ce critère porte le nom d'*erreur quadratique moyenne intégrée* (MISE)<sup>6</sup>.

La convergence en *moyenne quadratique intégrée* de l'estimateur  $\hat{f}_\Delta^n$  défini par (I.6) a été établie par Lecoutre [67]. Nous avons, d'après ([106] page 17) :

$$\text{MISE} = \frac{1}{n\Delta} + \frac{\Delta^2 R(f')}{12} + \mathcal{O}(n^{-1}) + \mathcal{O}(\Delta^3), \quad (\text{I.10})$$

où  $R(s) = \int_\Omega (s(x))^2 dx$ , pour toute fonction  $s$  de carré intégrable.

Cette erreur quadratique moyenne intégrée tend vers zéro quand  $\Delta$  tend vers zéro et  $n\Delta$  tend vers l'infini quand  $n$  tend vers l'infini.

L'utilisation du critère MISE permet de définir une *granularité optimale* de l'histogramme notée  $\Delta^*$ . Cette valeur optimale est la valeur qui minimise ce critère pour un nombre d'observations et une loi donnée. Cette valeur optimale est de la forme :

$$\Delta^* = \left[ \frac{6}{R(f')} \right]^{\frac{1}{3}} n^{-\frac{1}{3}}. \quad (\text{I.11})$$

Notons que cette valeur est inutilisable en pratique car elle fait intervenir une connaissance a priori de densité inconnue  $f$  via l'intégrale du carré de sa dérivée ( $R(f')$ ). Puisque, justement, cette technique est dédiée à l'estimation de densités dont la loi est inconnue, la valeur de  $\Delta^*$  est généralement inconnue.

En remplaçant, la valeur de  $\Delta^*$  (I.11) dans l'expression (I.10), on obtient la valeur optimale de *l'erreur quadratique moyenne intégrée asymptotique* (AMISE)<sup>7</sup> notée AMISE\* :

$$\text{AMISE}^* = \left[ \frac{9R(f')}{16} \right]^{\frac{1}{3}} n^{-\frac{2}{3}}. \quad (\text{I.12})$$

Geffroy a montré dans [44] la convergence *uniforme*<sup>8</sup> et *presque complète*<sup>9</sup> de l'estimateur  $\hat{f}_\Delta^n$  (I.6). Tandis que la *convergence en probabilité*<sup>10</sup> et presque complète en norme  $L^1(\Omega)$  a été établie par Abou-Jaoudé [1].

### I.2.1.5 Utilisation de l'estimateur de densité par histogrammes

Utiliser un histogramme pour estimer la densité présente l'intérêt de la simplicité. Cette simplicité a cependant une contre partie désavantageuse. En effet, comme expliqué dans ([14] pages 26-28), la représentation par histogramme de la densité sous-jacente à un ensemble fini d'observations n'est robuste au choix ni de la granularité ni de l'intervalle de référence de la partition sur laquelle l'histogramme est bâti.

Nous illustrons ce manque de robustesse sur un exemple de 300 observations tirées aléatoirement d'une distribution normale  $\mathcal{N}(0, 1)$ . Sur la Figure I.3, nous avons choisi deux valeurs de largeur de cellule :  $\Delta = 2.5$  (Figure I.3 (a)) et  $\Delta = 0.1$  (Figure I.3

6. En anglais Mean Integrated Squared Error.

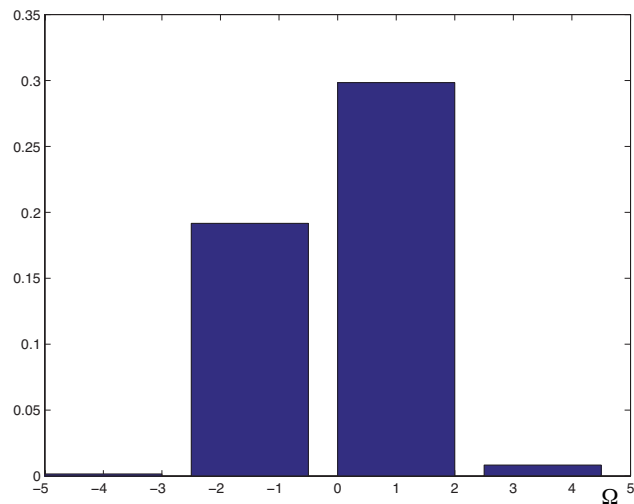
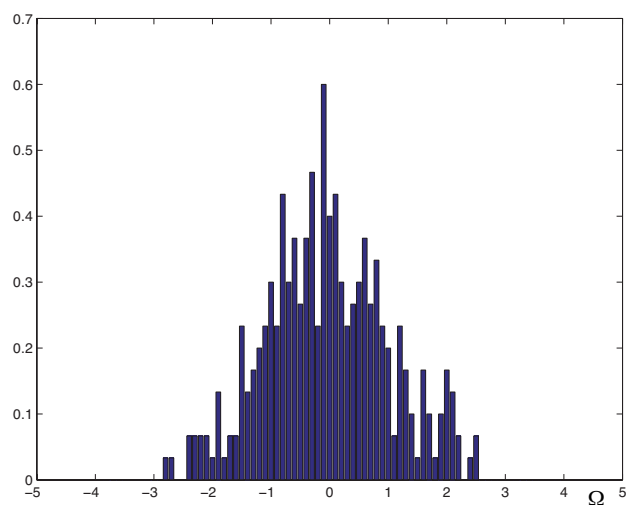
7. En anglais Asymptotic Mean Integrated Squared Error.

8. Un estimateur  $\hat{s}^n$  converge uniformément vers  $s \in L(\Omega)$  si  $\lim_{n \rightarrow \infty} \{\sup_{u \in \Omega} |\hat{s}^n(u) - s(u)|\} = 0$  ([12] page 82).

9. Un estimateur  $\hat{s}^n$  converge presque complète vers une fonction  $s \in L(\Omega)$  si pour tout  $u \in \Omega$  et tout  $\epsilon > 0$ , on a  $\sum_n P(|s_n(u) - s(u)| > \epsilon) < \infty$  ([4] page 110).

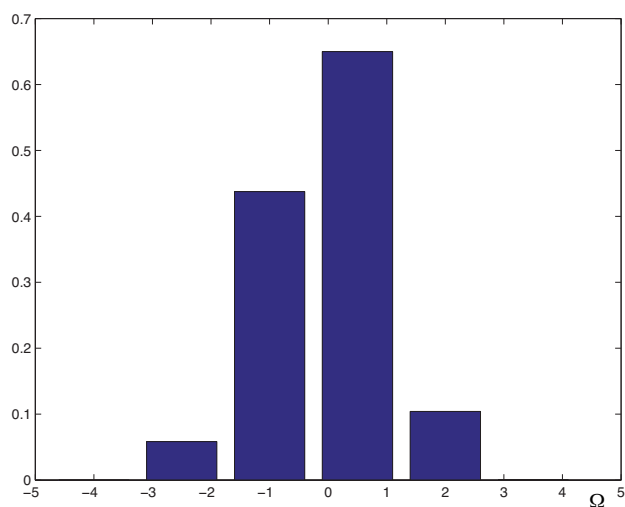
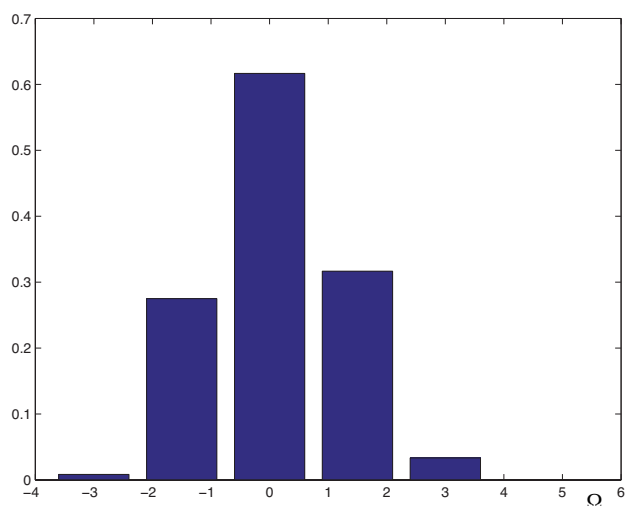
10. Un estimateur  $\hat{s}^n$  converge en probabilité vers  $s \in L(\Omega)$  si, pour tout  $u \in \Omega$  et tout  $\epsilon > 0$  on a  $\lim_{n \rightarrow \infty} P(|s_n(u) - s(u)| \geq \epsilon) = 0$  ([4] page 113).

(b)). On peut, par exemple, s'intéresser à la localisation du mode principal (c'est-à-dire la valeur la plus fréquemment prise par ces observations) sur l'histogramme. Pour  $\Delta = 2.5$ , la localisation du mode est claire dans l'intervalle  $[0, 2]$ . Tandis que pour  $\Delta = 0.1$ , la localisation de ce mode est dans l'intervalle  $[-0.14, -0.06]$ .

(a)  $\Delta = 2.5$ (b)  $\Delta = 0.1$ 

**Figure I.3** : Histogramme de 300 observations construit avec une granularité  $\Delta = 2.5$  (a) et  $\Delta = 0.1$  (b).

Sur la Figure I.4, nous avons choisi deux positions différentes de l'intervalle de référence  $\Omega$  :  $\Omega = [-5, 5]$  (Figure I.4 (a)) et  $\Omega = [-4, 6]$  (Figure I.4 (b)). La Figure I.4 montre l'influence du positionnement de  $\Omega$ , sur la représentation d'observations. Sur la Figure I.4 (b), la distribution semble symétrique alors qu'elle semble ne pas l'être sur la Figure I.4 (a).

(a)  $\Omega = [-5, 5]$ (b)  $\Omega = [-4, 6]$ 

**Figure I.4 :** Histogramme de 300 observations construit sur la portion  $\Omega = [-5, 5]$  (a) et  $\Omega = [-4, 6]$  (b).

Runkler [95], Van Den Berg [116] et Strauss et al. [110] ont tenté de réduire l'influence du partitionnement arbitraire de l'intervalle de référence  $\Omega$  en remplaçant le partitionnement binaire par un *partitionnement flou* de  $\Omega$ . L'estimateur de densité par des histogrammes basé sur un partitionnement flou ainsi que ses propriétés statistiques ont été établies par Waltman et al. [122] et Loquin et Strauss [70].

Du point de vue de l'estimation non-paramétrique de densité, cet estimateur est une fonction étagée, et donc discontinue. À cause de cette discontinuité, l'histogramme ne peut pas être adapté à la situation où nous disposons d'une information à priori sur la régularité de la densité à estimer. Par exemple, si l'on sait par avance que la densité de probabilité est deux fois continûment différentiable, l'estimation naturelle de cette densité doit être deux fois continûment différentiable. Les histogrammes, qui produisent des estimations

discontinues, ne peuvent donc pas répondre à ce type de problématique.

### I.2.2 Estimateur simple de densité

L'estimateur simple de densité, appelé aussi la méthode d'estimation par les histogrammes mobiles, a été proposé par Fix et Hodges [43]. L'estimation de la densité en un point  $x \in \Omega$ , par cette méthode, consiste à construire autour de  $x$  un intervalle (ou *fenêtre*) de largeur  $\Delta$  ( $[x - \frac{\Delta}{2}, x + \frac{\Delta}{2}]$ ) et à compter le nombre d'observations dans cet intervalle.

Partons du lien existant entre la densité de probabilité  $f$  et la fonction de répartition  $F$  :

$$\forall x \in \Omega, F(x) = \int_{-\infty}^x f(u) du, \quad (\text{I.13})$$

on peut écrire

$$\begin{aligned} f(x) &= \lim_{\Delta \rightarrow 0} \frac{P(x - \frac{\Delta}{2} \leq x_i \leq x + \frac{\Delta}{2})}{\Delta} \\ &= \lim_{\Delta \rightarrow 0} \frac{F(x + \frac{\Delta}{2}) - F(x - \frac{\Delta}{2})}{\Delta}. \end{aligned}$$

L'estimateur simple de  $f$ , noté  $\hat{f}_{\Delta}^n$ , peut alors être défini, pour tout  $x \in \Omega$ , par :

$$\begin{aligned} \hat{f}_{\Delta}^n(x) &= \frac{1}{\Delta} \frac{|\{i : x - \frac{\Delta}{2} \leq x_i \leq x + \frac{\Delta}{2}\}|}{n} \\ &= \frac{1}{n\Delta} \sum_{i=1}^n \mathbb{1}_{[x - \frac{\Delta}{2}, x + \frac{\Delta}{2}]}(x_i) \\ &= \frac{1}{n\Delta} \sum_{i=1}^n \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(\frac{x - x_i}{\Delta}), \end{aligned} \quad (\text{I.14})$$

où  $|\cdot|$  est le cardinal d'un ensemble. Cet estimateur peut aussi s'écrire

$$\hat{f}_{\Delta}^n(x) = \frac{1}{n\Delta} \sum_{i=1}^n \mathfrak{w}\left(\frac{x - x_i}{\Delta}\right), \quad (\text{I.15})$$

où  $\mathfrak{w}$  est une *fonction de poids* définie par :

$$\mathfrak{w}(x) = \begin{cases} 1 & \text{si } x \in [-\frac{1}{2}, \frac{1}{2}], \\ 0 & \text{sinon.} \end{cases} \quad (\text{I.16})$$

Cette fonction de poids n'est autre que la densité de probabilité uniforme sur l'intervalle  $[-\frac{1}{2}, \frac{1}{2}]$ .

Le biais de l'estimateur  $\hat{f}_{\Delta}^n$  (I.15) est donné ([93] page 371), pour tout  $x \in \Omega$ , par :

$$\begin{aligned} \text{biais}(\hat{f}_{\Delta}^n(x)) &= \mathbb{E}(\hat{f}_{\Delta}^n(x)) - f(x) \\ &= \frac{1}{\Delta} (F(x + \frac{\Delta}{2}) - F(x - \frac{\Delta}{2})) - f(x). \end{aligned} \quad (\text{I.17})$$

L'expression (I.17) montre que ce biais converge vers zéro quand la largeur de fenêtre  $\Delta$  tend vers zéro.



La variance de cet estimateur est donnée ([93] page 371), pour tout  $x \in \Omega$ , par :

$$\text{var}(\hat{f}_\Delta^n(x)) = \frac{1}{2n\Delta^2} \left( F(x + \frac{\Delta}{2}) - F(x - \frac{\Delta}{2}) \right) \left( 1 - F(x + \frac{\Delta}{2}) + F(x - \frac{\Delta}{2}) \right). \quad (\text{I.18})$$

Cette variance tend vers zéro lorsque  $n\Delta$  tend vers l'infini quand le nombre d'observations  $n$  tend vers l'infini. Ce qui donne ([93] page 371) la convergence en moyenne quadratique :

$$\text{si } \Delta \rightarrow 0 \text{ et } n\Delta \rightarrow \infty \text{ quand } n \rightarrow \infty, \text{MSE}(\hat{f}_\Delta^n(x)) = \mathbb{E}(\hat{f}_\Delta^n(x) - f(x))^2 \rightarrow 0. \quad (\text{I.19})$$

L'inconvénient majeur de l'estimateur simple défini par l'expression (I.15) est comme l'estimation par histogramme de fournir une estimation discontinue. Elle est cependant discontinue uniquement aux points  $\{(x_i - \frac{\Delta}{2}, x_i + \frac{\Delta}{2})\}_{i \in \{1, \dots, n\}}$  contrairement à l'estimateur par histogramme qui est discontinue aux frontières de chaque cellule et a une dérivée nulle partout ailleurs. Cette discontinuité est une conséquence de la discontinuité de la fonction de poids  $\mathbf{w}$  (I.16).

Comme le souligne Silverman ([104] pages 21-22), cette méthode d'estimation est finalement comparable à celle de l'histogramme. En effet, considérons un histogramme construit à partir d'intervalles de largeur  $\Delta$ . Si on suppose qu'aucune des observations ne soit aux bornes de ces intervalles et si  $x \in \Omega$  est au centre de l'un de ces intervalles alors l'estimateur simple, au point  $x$ ,  $\hat{f}_\Delta^n(x)$  (I.15) n'est autre que l'estimateur par histogramme  $\hat{f}_\Delta^n(x)$  (I.6). L'estimateur simple peut être vu comme une manière de construire un histogramme où chaque point est le centre d'un intervalle de l'ensemble  $\{A_k, k = 1, \dots, p\}$  de l'histogramme.

### I.2.3 Estimateur à noyau de la densité

L'estimateur simple de la densité défini par l'expression (I.15) peut être généralisé en remplaçant la fonction de poids  $\mathbf{w}$  (I.16) (qui est une densité de probabilité uniforme) par une fonction de poids plus générale, notée  $\kappa$ , (qui est une densité de probabilité quelconque). Cette nouvelle fonction de poids est appelée *noyau sommatif* [72] (ou simplement *noyau*).

**Définition I.8 (Noyau sommatif [72])** *Un noyau sommatif est une fonction  $\kappa : \Omega \rightarrow \mathbb{R}^+$ , qui vérifie la propriété de sommativité, c'est-à-dire :*

$$\int_{\Omega} \kappa(u) du = 1. \quad (\text{I.20})$$

On peut définir, à partir de noyau sommatif  $\kappa$ , un noyau sommatif  $\kappa_\Delta^x$  translaté en  $x \in \Omega$  et dilaté d'une *largeur de bande (ou paramètre de lissage)*  $\Delta > 0$ , par :

$$\forall u \in \Omega, \kappa_\Delta^x(u) = \frac{1}{\Delta} \kappa\left(\frac{u-x}{\Delta}\right). \quad (\text{I.21})$$

Par convention,  $\kappa(u) = \kappa_1^0(u)$ .

La Figure I.5 présente quatre noyaux sommatifs ayant une largeur de bande  $\Delta = 1$ . L'intervalle de référence est  $\Omega = [-2, 2]$  (voir, par exemple, ([106] page 44) pour la forme analytique de chacun de ces noyaux sommatifs).

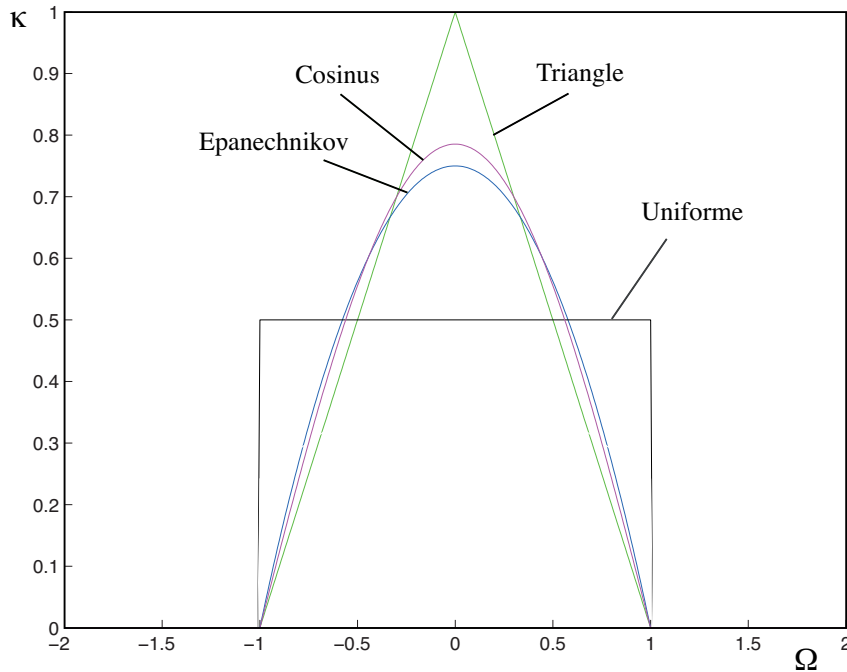


Figure I.5 : Superposition de quatre noyaux sommatifs.

La plupart des noyaux sommatifs couramment utilisés en estimation fonctionnelle sont monomodaux, symétriques et centrés (c'est-à-dire définissant un voisinage autour de l'origine). Dans la suite de ce manuscrit, nous appelons  $\mathcal{K}(\Omega)$  l'ensemble des noyaux sommatifs monomodaux, symétriques et centrés.

Un noyau sommatif  $\kappa$  peut être vu comme une densité de probabilité induisant une mesure de probabilité, notée  $P_\kappa$ , définie par :

$$\forall A \subseteq \Omega, P_\kappa(A) = \int_A \kappa(u) du. \quad (\text{I.22})$$

Sous cet angle, on peut définir l'opérateur *espérance sommative* basée sur un noyau sommatif  $\kappa$  :

**Définition I.9 (Espérance sommative [72])** Soit  $s$  une fonction de  $L(\Omega)$  et soit  $\kappa$  un noyau sommatif de  $\mathcal{K}(\Omega)$ , l'espérance sommative de  $s$  dont le voisinage est défini par  $\kappa$  est l'espérance mathématique de  $s$  par rapport à  $\kappa$  :

$$\mathbb{E}_\kappa(s) = \int_\Omega s dP_\kappa = \int_\Omega s(u) \kappa(u) du. \quad (\text{I.23})$$

### I.2.3.1 Principe et définition

L'estimateur à noyau, appelé aussi estimateur de densité de Parzen-Rosenblatt a été introduit par Rosenblatt [94] et développé par Parzen [85].

**Définition I.10 (Estimateur de densité de Parzen-Rosenblatt [85, 94])** Soit  $f$  une densité de probabilité sous-jacente à un ensemble fini d'observations  $(x_1, \dots, x_n)$ . L'estimateur de densité à noyau de Parzen-Rosenblatt, noté  $\hat{f}_{\kappa, \Delta}^n$ , est défini, pour tout  $x \in \Omega$ ,

par :

$$\widehat{f}_{\kappa\Delta}^n(x) = \frac{1}{n\Delta} \sum_{i=1}^n \kappa\left(\frac{x-x_i}{\Delta}\right), \quad (\text{I.24})$$

où  $\kappa$  est un noyau sommatif (Définition I.8).

L'estimateur de densité de Parzen-Rosenblatt  $\widehat{f}_{\kappa\Delta}^n(x)$  (I.24) tend à estimer la densité en chaque point  $x \in \Omega$ , en centrant d'abord un noyau  $\kappa$  sur chaque observation  $x_i \in \Omega$ . Ensuite, chaque observation  $x_i$  contribue à la densité  $f$  suivant sa distance par rapport à  $x$  et l'estimateur  $\widehat{f}_{\kappa\Delta}^n(x)$  au point  $x$  est proportionnel à la somme des contributions de tous les  $x_i$ .

L'estimateur de densité de Parzen-Rosenblatt  $\widehat{f}_{\kappa\Delta}^n$  est l'estimateur le plus populaire parmi les multiples estimateurs d'estimation non-paramétrique de la densité [26, 99]. Comme le font remarquer Berlinet et Biau ([7] page 26), le succès rencontré par cet estimateur peut être résumé en trois points. Le premier point est que l'estimateur de densité de Parzen-Rosenblatt, donné par son expression (I.24), est simple puisque  $\widehat{f}_{\kappa\Delta}^n$  est la somme de  $n$  variables aléatoires indépendantes et identiquement distribuées. Le deuxième point porte sur le fait que l'estimateur  $\widehat{f}_{\kappa\Delta}^n$  (I.24) converge vers la vraie densité  $f$  en  $L^1(\Omega)$  quand  $\Delta$  tend vers zéro et  $n\Delta$  tend vers l'infini quand  $n$  tend vers l'infini. La convergence de cet estimateur a été établie dans tous les modes, c'est-à-dire en probabilité, en moyenne quadratique, presque sûrement et presque complètement. Le troisième point est la flexibilité de l'estimateur à noyau (I.24), dans le sens où il laisse à l'utilisateur une grande latitude non seulement dans le choix du noyau sommatif  $\kappa$ , mais aussi dans le choix de la largeur de bande  $\Delta$ .

### I.2.3.2 Propriétés statistiques de l'estimateur à noyau

Nous présentons, dans cette partie, les propriétés statistiques de l'estimateur de densité de Parzen-Rosenblatt défini par l'expression (I.24) (pour une étude plus détaillée, voir par exemple Bosq et Lecoutre [10], Tsybakov [115] ou Silverman [104]).

Nous considérons ici les noyaux sommatifs  $\kappa$  centrés de variance finie de  $\mathcal{K}(\Omega)$  c'est-à-dire vérifiant :

$$\int_{\Omega} u\kappa(u)du = 0 \text{ et } \int_{\Omega} u^2\kappa(u)du = \sigma_{\kappa}^2 > 0, \quad (\text{I.25})$$

$\sigma_{\kappa}^2$  est appelée la *variance du noyau*  $\kappa$  [7].

Le biais de l'estimateur  $\widehat{f}_{\kappa\Delta}^n$  (I.24) est donné ([106] page 42), pour tout  $x \in \Omega$ , par :

$$\text{biais}(\widehat{f}_{\kappa\Delta}^n(x)) = \frac{\Delta^2\sigma_{\kappa}^2 f''(x)}{2} + \mathcal{O}(\Delta^4), \quad (\text{I.26})$$

où  $f''$  est la dérivée seconde de la fonction de densité  $f$ .  $f''$  doit être une fonction absolument continue.

L'expression (I.26) montre que ce biais converge vers zéro quand la largeur de bande  $\Delta$  tend vers zéro. Il découle de l'expression (I.26) que le biais de  $\widehat{f}_{\kappa\Delta}^n$  ne dépend pas directement du nombre d'observations mais seulement de la fonction de noyau  $\kappa$  et donc l'augmentation du nombre d'observations ne peut pas, à elle seule, diminuer le biais.

La variance de cet estimateur est donnée ([106] page 42), pour tout  $x \in \Omega$ , par :

$$\text{var}(\widehat{f}_{\kappa_\Delta}^n(x)) = \frac{f(x)R(\kappa)}{n\Delta} + \mathcal{O}(n^{-1}), \quad (\text{I.27})$$

où  $R(\kappa) = \int_{\Omega} (\kappa(u))^2 du$ .

Cette variance tend vers zéro quand  $n\Delta$  tend vers l'infini quand  $n$  tend vers l'infini.

La convergence en moyenne quadratique de l'estimateur  $\widehat{f}_{\kappa_\Delta}^n$  (I.24) a été établie par Bosq et Lecoutre [10]. Nous avons, d'après ([106] page 42), pour tout  $x \in \Omega$  :

$$\text{MSE}(\widehat{f}_{\kappa_\Delta}^n(x)) = \frac{f(x)R(\kappa)}{n\Delta} + \frac{\Delta^4 \sigma_\kappa^4 (f''(x))^2}{4} + \mathcal{O}(n^{-1}) + \mathcal{O}(\Delta^6). \quad (\text{I.28})$$

En intégrant l'expression (I.28) sur tout le domaine  $\Omega$ , on obtient ([106] page 42) l'erreur quadratique moyenne intégrée de  $\widehat{f}_{\kappa_\Delta}^n$  :

$$\text{MISE} = \frac{R(\kappa)}{n\Delta} + \frac{\Delta^4 \sigma_\kappa^4 R(f'')}{4} + \mathcal{O}(n^{-1}) + \mathcal{O}(\Delta^6), \quad (\text{I.29})$$

où  $R(f'') = \int_{\Omega} (f''(u))^2 du$ .

La largeur de bande optimale de  $\widehat{f}_{\kappa_\Delta}^n$ , notée  $\Delta^*$ , est donnée ([106] page 43) par :

$$\Delta^* = \left( \frac{R(\kappa)}{\sigma_\kappa^4 R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \quad (\text{I.30})$$

Notons que  $\Delta^*$  est une quantité déterministe qui dépend du nombre d'observations  $n$ . Elle dépend aussi de la densité inconnue  $f$  via l'intégrale du carré de sa dérivée seconde ( $R(f'')$ ) et ne peut pas donc être utilisée telle qu'elle dans les calculs. Une façon classique de corriger ce défaut consiste à remplacer la quantité  $R(f'')$ , dans l'expression (I.30), par un estimateur approprié [7]. Cette correction a fait l'objet de beaucoup de travaux (voir, par exemple, Hall et Marron [51], Jones et Sheather [57]).

En remplaçant la valeur de  $\Delta^*$  (I.30) dans l'expression (I.29), on obtient ([106] page 44) la valeur du AMISE\* :

$$\text{AMISE}^* = \frac{5}{4} (\sigma_\kappa R(\kappa))^{\frac{4}{5}} (R(f''))^{\frac{1}{5}} n^{-\frac{4}{5}}. \quad (\text{I.31})$$

Epanechnikov a proposé dans [40] un *noyau optimal* de l'estimateur  $\widehat{f}_{\kappa_\Delta}^n(x)$  (I.24) qui minimise le critère AMISE\* (I.31) par rapport au noyau  $\kappa$ . La solution  $\kappa_{opt}$  obtenue est donnée par :

$$\kappa_{opt}(u) = \begin{cases} \frac{3}{4}(1-u^2) & \text{si } |u| \leq 1, \\ 0 & \text{sinon.} \end{cases} \quad (\text{I.32})$$

Ce noyau est appelé le *noyau d'Epanechnikov*. Nous notons, dans la suite,  $\mathfrak{e}$  et non  $\kappa_{opt}$  le noyau d'Epanechnikov défini par (I.32).

Le couple largeur de bande  $\Delta$  et forme de noyau  $\kappa$  joue un rôle important dans la méthode d'estimation à noyau. Nous discutons, dans la section suivante, de l'influence de ce choix sur l'estimation de la densité  $f$  sous-jacente à un ensemble fini d'observations.

La convergence ponctuelle presque sûre de l'estimateur  $\widehat{f}_{\kappa_\Delta}^n$  (I.24) a été établie par Devroy et Wagner [29]. Bertrand Retali a montré dans [91] la convergence uniforme presque

sûre de cet estimateur, tandis que la convergence en  $L^1(\Omega)$  a été établie par Gilck [48] et Devroy [26].

Un résultat important issu de [104] montre que si  $f''$ , la dérivée seconde de  $f$ , existe et que la largeur de bande  $\Delta = qn^{-\frac{1}{5}}$  (avec  $q$  une constante), alors, pour tout  $x \in \Omega$ , l'estimateur à noyau  $\hat{f}_{\kappa\Delta}^n(x)$  est asymptotiquement normal (c'est-à-dire converge en loi<sup>11</sup> vers la loi normale) :

$$n^{\frac{2}{5}} \{ \hat{f}_{\kappa\Delta}^n(x) - f(x) \} \xrightarrow[n \rightarrow \infty]{L} \mathcal{N} \left( \frac{q^2}{2} f''(x) \sigma_\kappa^2, \frac{1}{q} f(x) R(\kappa) \right), \quad (\text{I.33})$$

où  $\sigma_\kappa^2$  est la variance de  $\kappa$  définie dans (I.25).

L'expression (I.33) est équivalente, par une transformation linéaire de la loi normale, à

$$\{ \hat{f}_{\kappa\Delta}^n(x) - f(x) \} \xrightarrow[n \rightarrow \infty]{L} \mathcal{N} \left( \underbrace{\frac{\Delta^2}{2} f''(x) \sigma_\kappa^2}_{\text{biais de } \hat{f}_{\kappa\Delta}^n}, \underbrace{\frac{1}{n\Delta} f(x) R(\kappa)}_{\text{variance de } \hat{f}_{\kappa\Delta}^n} \right). \quad (\text{I.34})$$

Sous l'hypothèse que le terme de biais (I.26) de  $\hat{f}_{\kappa\Delta}^n$  est négligeable par rapport au terme de la variance (I.27) de  $\hat{f}_{\kappa\Delta}^n$ , l'*intervalle de confiance* au seuil  $\alpha$  sur la valeur de la densité  $f(x)$ , pour tout  $x \in \Omega$ , est donné ([52] page 62) par :

$$\left[ \hat{f}_{\kappa\Delta}^n(x) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{f}_{\kappa\Delta}^n(x) R(\kappa)}{n\Delta}}, \hat{f}_{\kappa\Delta}^n(x) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{f}_{\kappa\Delta}^n(x) R(\kappa)}{n\Delta}} \right], \quad (\text{I.35})$$

où  $z_{1-\frac{\alpha}{2}}$  est le  $(1 - \frac{\alpha}{2})$ -quantile<sup>12</sup> de la loi normale centrée réduite.

Notons que cette méthode donne un intervalle de confiance à  $(1 - \frac{\alpha}{2})\%$  des valeurs de la densité  $f$  et non pas pour toutes les valeurs de  $f$ . Une autre technique de calcul proposée par Bickel et Rosenblatt [8] sous le nom de *bandes de confiance*,<sup>13</sup> permet de donner un autre intervalle d'estimation de la densité  $f$ . Sous l'hypothèse que  $f$  est une densité sur  $[0, 1]$  et que  $\Delta = n^{-l}$ ,  $l \in (\frac{1}{5}, \frac{1}{2})$ , Bickel et Rosenblatt ont démontré que, pour tout  $x \in [0, 1]$ , nous avons le résultat asymptotique suivant :

$$\lim_{n \rightarrow \infty} P \left( \hat{f}_{\kappa\Delta}^n(x) - Z_{n,\alpha} \sqrt{\frac{\hat{f}_{\kappa\Delta}^n(x) R(\kappa)}{n\Delta}} \leq f(x) \leq \hat{f}_{\kappa\Delta}^n(x) + Z_{n,\alpha} \sqrt{\frac{\hat{f}_{\kappa\Delta}^n(x) R(\kappa)}{n\Delta}} \right) = 1 - \alpha. \quad (\text{I.36})$$

où

$$Z_{n,\alpha} = \left( - \frac{\log\{-\frac{1}{2} \log(1 - \alpha)\}}{(2l \log n)^{\frac{1}{2}}} + d_n \right)^{\frac{1}{2}}, \quad (\text{I.37})$$

avec

$$d_n = (2l \log n)^{\frac{1}{2}} + (2l \log n)^{-\frac{1}{2}} \log \left( \frac{1}{2\pi} \frac{R(\kappa')}{R(\kappa)} \right)^{\frac{1}{2}}, \quad (\text{I.38})$$

11. Une suite des variables aléatoires  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers la variable aléatoire  $X$  si, pour tout  $x \in \Omega$ , on a :  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ , où  $F_X$  est la fonction de répartition de  $X$ . On le note  $X_n \xrightarrow[n \rightarrow \infty]{L} X$  ([68] pages 79).

12. Si  $X$  est une variable aléatoire suivant la loi normale centrée réduite, alors la valeur du quantile d'ordre  $1 - \frac{\alpha}{2}$  c'est-à-dire la valeur  $z_{1-\frac{\alpha}{2}}$  telle que  $P(|X| \geq z_{1-\frac{\alpha}{2}}) = \alpha$  ([68] page 58).

13. En anglais confidence bands.

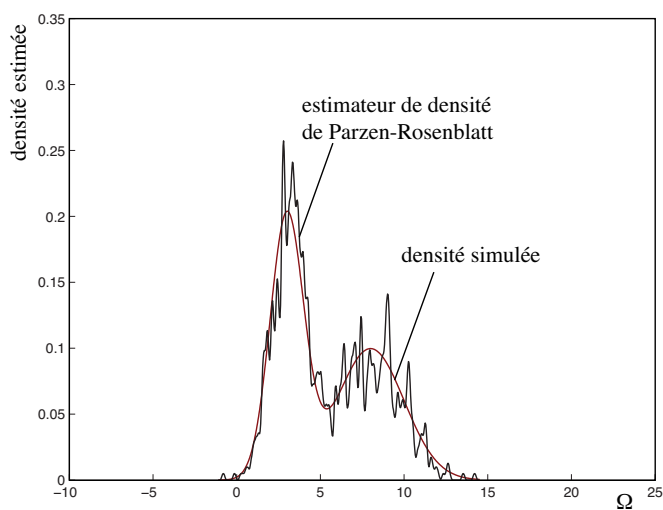
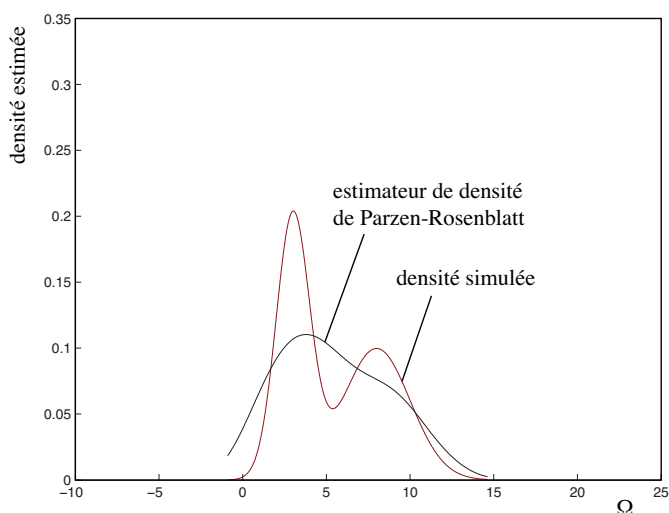
où  $\kappa'$  est la dérivée de noyau  $\kappa$  et  $R(\kappa') = \int_{\Omega} (\kappa'(u))^2 du$ .

Comme le fait remarquer Härdle ([52] pages 66-67) les intervalles d'estimation de  $f$  donnés par les bandes de confiance (I.36) sont *plus étroits* que ceux donnés par les intervalles de confiance (I.35).

### I.2.3.3 Utilisation de l'estimateur de densité par noyau

L'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_{\kappa\Delta}^n$  défini par l'expression (I.24) dépend de deux paramètres : la forme fonctionnelle du noyau  $\kappa$  et la largeur de bande  $\Delta$ . De manière générale, il est admis le choix du noyau  $\kappa$  a beaucoup moins d'importance que celui de la largeur de bande  $\Delta$ .  $\Delta$  détermine le degré de lissage de l'estimation d'une densité. Une faible largeur de bande implique un faible degré de lissage et résulte en une fonction de densité irrégulière. Une large valeur de  $\Delta$  conduit à une estimation lisse. La Figure I.6 présente le problème du choix de largeur de bande.

Sur la Figure I.6, nous avons choisi deux valeurs différentes de  $\Delta$  :  $\Delta = 0.2$  (Figure I.6 (a)) et  $\Delta = 5$  (Figure I.6 (b)). Dans cet exemple, nous avons simulé 1000 observations d'une variable aléatoire dont la densité de probabilité est une distribution bimodale obtenue en contaminant une loi normale  $\mathcal{N}(3, 1)$  par une autre  $\mathcal{N}(8, 4)$ . L'intervalle de référence est  $\Omega = [-10, 25]$ . Nous avons choisi, pour réaliser l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_{\kappa\Delta}^n$  (I.24), le noyau cosinus comme noyau sommatif de référence. Sur les deux Figures I.6 (a et b), la densité de probabilité simulée  $f$  est dessinée en rouge et l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_{\kappa\Delta}^n$  en noir. On peut voir l'influence de la largeur de bande sur l'estimation de la densité. Pour  $\Delta = 0.2$  la densité de probabilité est sous-lissée et l'aspect bimodal de  $f$  n'est pas visible, pas plus que dans le cas  $\Delta = 5$ , où l'estimation de la densité de probabilité est sur-lissée et aplatie.

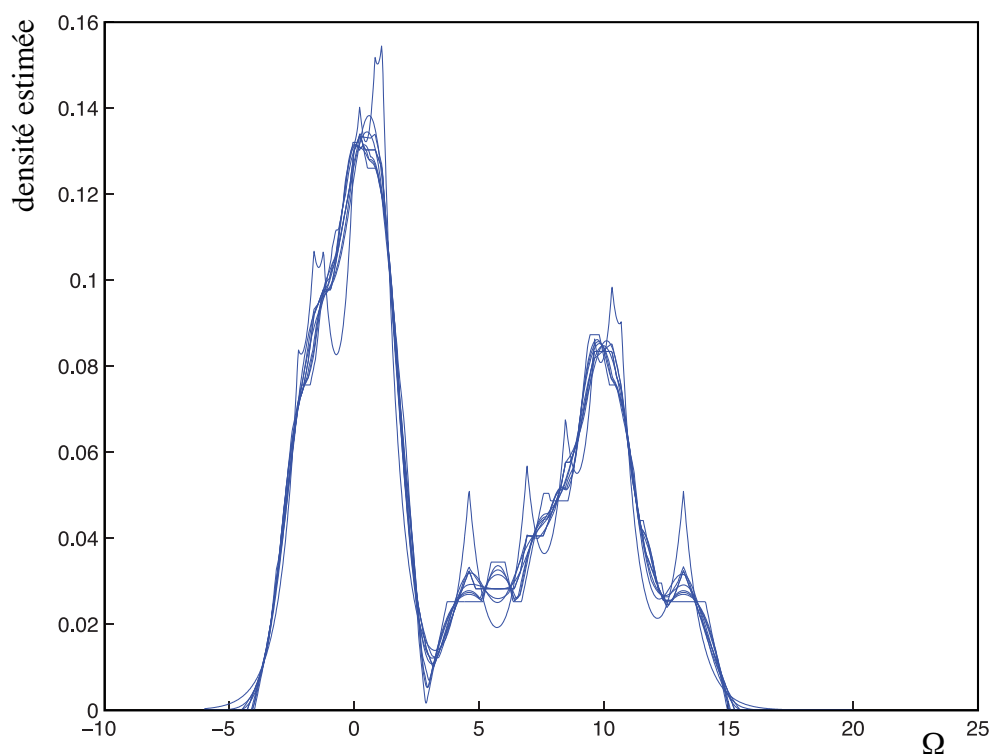
(a)  $\Delta = 0.2$ (b)  $\Delta = 5$ 

**Figure I.6 :** L'estimateur de densité de Parzen-Rosenblatt construit à partir de 1000 observations en utilisant le noyau cosinus avec une largeur de bande  $\Delta = 0.2$  (a) et  $\Delta = 5$  (b).

Rao a souligné dans [88] que le choix du noyau  $\kappa$  autre que le noyau optimal (noyau d'Epanechnikov) (I.32) ne menait qu'à une faible perte de précision. Lall et al. ont considéré dans [66] que le choix du noyau a une certaine importance, mais que son influence sur l'ensemble de l'estimation est relativement faible. Généralement, les arguments utilisés sur le choix du noyau sont basés sur le comportement asymptotique quand on dispose d'un grand nombre d'observations. Cependant, lorsque le nombre d'observations disponible n'est pas grand, le choix de la forme du noyau sommatif  $\kappa$  a une influence certaine.

La Figure I.7 présente une expérience illustrant l'influence de la forme de noyau sommatif sur l'estimation de la densité quand le nombre d'observations  $n$  est faible. Sur la Figure I.7, nous avons considéré un ensemble de 30 observations issues de la même loi

bimodale simulée (Figure I.6) en utilisant 9 noyaux sommatifs différents avec une largeur de bande  $\Delta$  adaptée à cette densité et à ce nombre d'observations ( $\Delta = 0.8$ ).



**Figure I.7 :** Variation sur la forme du noyau  $\kappa$  pour la méthode d'estimation par noyau.

L'examen de la Figure I.7 montre l'influence de la variation sur la forme du noyau  $\kappa$  sur l'estimation de la densité. La *méthode d'estimation imprécise de la densité*, que nous proposons, permet de calculer la borne inférieure et la borne supérieure d'un ensemble d'estimations obtenues en considérant des noyaux sommatifs appartenant à un ensemble fini convexe des noyaux.

## I.3 Estimation non-paramétrique de la fonction de répartition

Nous présentons, dans cette section, deux méthodes d'estimation non-paramétriques de la fonction de répartition  $F$  ainsi que les propriétés statistiques de chaque méthode d'estimation (pour une étude bien détaillée sur cette problématique voir, par exemple, ([101] pages 87-108)).

### I.3.1 Estimation naturelle de la fonction de répartition par la fonction de répartition empirique

La *fonction de répartition empirique* est un estimateur simple et performant de la fonction de répartition. Soit  $(x_1, \dots, x_n)$  un ensemble fini d'observations de  $n$  variables aléatoires  $(X_1, \dots, X_n)$  i.i.d de fonction de répartition commune  $F$  inconnue. La fonction



de répartition empirique, notée  $E_n$ , est définie, pour tout  $x \in \Omega$ , par :

$$E_n(x) = \frac{1}{n} \sum_{i=1}^n H(x - x_i), \quad (\text{I.39})$$

où  $H$  est la *fonction de Heaviside* définie par :

$$H(x) = \begin{cases} 1 & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases} \quad (\text{I.40})$$

L'estimateur  $E_n$  est, d'après Yamato [126], l'unique estimateur parmi les estimateurs sans biais<sup>14</sup> de variance minimum.

L'espérance de l'estimateur  $E_n$  (I.39) est donnée, pour tout  $x \in \Omega$ , par :

$$\mathbb{E}(E_n(x)) = F(x). \quad (\text{I.41})$$

La variance de  $E_n$  est donnée, pour tout  $x \in \Omega$ , par :

$$\text{var}(E_n(x)) = \frac{F(x)(1 - F(x))}{n}. \quad (\text{I.42})$$

En se basant sur ces expressions, pour tout  $x \in \Omega$ , la convergence en moyenne quadratique de  $E_n(x)$  vers  $F(x)$  :

$$\text{quand } n \longrightarrow \infty, \text{MSE}(E_n(x)) = \mathbb{E}(E_n(x) - F(x))^2 \longrightarrow 0. \quad (\text{I.43})$$

Dans son article, Yamato a démontré la convergence de la distribution :

$$\frac{\sqrt{n}(E_n(x) - \mathbb{E}(F(x)))}{\sqrt{F(x)(1 - F(x))}},$$

vers une loi normale  $\mathcal{N}(0, 1)$ .

Dans le cadre d'étude de l'estimateur  $E_n$  (I.39), nous définissons les statistiques suivantes :

$$S_n^- = \sup_{x \in \Omega} (-T_n(x)), \quad S_n^+ = \sup_{x \in \Omega} T_n(x), \quad \text{et } S_n = \sup_{x \in \Omega} |T_n(x)|, \quad (\text{I.44})$$

où  $\forall x \in \Omega, T_n(x) = \sqrt{n}(E_n(x) - F(x))$ . Kolmokorov [61] et Smirnov [108] ont montré que ces trois statistiques et leurs distributions ne dépendent pas de  $F$ . D'autre part, Donsker [30] et Doob [31] ont prouvé que les statistiques  $S_n^-$  et  $S_n^+$  ont la même loi et ont montré, pour tout constante  $\beta > 0$ , les résultats asymptotiques suivants :

$$\lim_{n \rightarrow \infty} P(S_n^- > \beta) = e^{-2\beta^2}, \quad (\text{I.45})$$

et

$$\lim_{n \rightarrow \infty} P(S_n^+ > \beta) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2\beta^2}. \quad (\text{I.46})$$

---

14. Un estimateur  $\hat{s}^n$  de  $s$  est dit estimateur sans biais si  $\mathbb{E}(\hat{s}^n) = s$ . C'est-à-dire  $\forall u \in \Omega, \mathbb{E}(\hat{s}^n(u)) - s(u) = 0$  ([53] page 122).

Grenander [50] a proposé un autre estimateur à partir de  $E_n$  (I.39). Cet estimateur, noté  $F_{Gr}$ , est défini comme le plus petit majorant concave<sup>15</sup> de  $E_n$ . Pour toute fonction  $E_n$  de support  $\Omega = [e_{min}, e_{max}]$ , la fonction  $F_{Gr}$  est la plus petite fonction concave telle que, pour tout  $x \in \Omega$ , on a :

$$\begin{cases} F_{Gr}(x) \geq E_n(x) \\ \text{et} \\ F_{Gr}(e_{min}) = 0, F_{Gr}(e_{max}) = 1. \end{cases} \quad (\text{I.47})$$

Cet estimateur est souvent utilisé, via sa dérivée, pour estimer la densité de probabilité  $f$  ([101] page 93).

Enfin, notons que la fonction de répartition empirique  $E_n$  définie par l'expression (I.39) est une fonction en escalier ne permettant pas de représenter la fonction de répartition  $F$  lorsque celle-ci est lisse.

### I.3.2 Estimation par noyau de la fonction de répartition

Rappelons d'abord que l'estimateur à noyau de la densité de probabilité  $f$  est donné (I.10), pour tout  $x \in \Omega$ , par :

$$\widehat{f}_{\kappa_{\Delta}}^n(x) = \frac{1}{n\Delta} \sum_{i=1}^n \kappa\left(\frac{x - x_i}{\Delta}\right),$$

où  $\kappa$  est un noyau sommatif (Définition I.8), et  $\Delta$  une largeur de bande.

L'estimateur à noyau de la fonction de répartition est alors défini, pour tout  $x \in \Omega$ , par :

$$\widehat{F}_{\kappa_{\Delta}}^n(x) = \frac{1}{n} \sum_{i=1}^n \kappa\left(\frac{x - x_i}{\Delta}\right), \quad (\text{I.48})$$

où  $\kappa(x) = \int_{-\infty}^x \kappa(u) du$ .

La convergence uniforme de l'estimateur à noyau  $\widehat{F}_{\kappa_{\Delta}}^n$  a été établie par Nadaraya [81] et Yamato [126] à condition que la fonction de densité  $f$  soit continue. Singh et al. ont démontré dans [107] cette convergence sans condition sur  $f$ .

La normalité asymptotique de l'estimateur  $\widehat{F}_{\kappa_{\Delta}}^n$  (I.48) a été prouvée par Watson et Leadbetter [123]. Winter a prouvé dans [124] que  $\widehat{F}_{\kappa_{\Delta}}^n$  vérifie, avec une probabilité égale à 1, la propriété de Chung-Smirnov c'est-à-dire :

$$\limsup_{n \rightarrow \infty} \left( \left( \frac{2n}{\log \log n} \right)^{\frac{1}{2}} \sup_{x \in \Omega} \left| \widehat{F}_{\kappa_{\Delta}}^n(x) - F(x) \right| \right) \leq 1. \quad (\text{I.49})$$

Azzalini a donné dans [3] l'expression asymptotique de l'erreur quadratique moyenne (MSE) ainsi que la largeur de bande optimale permettant d'avoir un MSE plus faible que celui de la fonction de répartition empirique  $E_n$  (I.39). Reiss a démontré dans [90] que

---

15. Soit  $s$  une fonction de  $L(\Omega)$  et soit  $l$  une constante telle que  $0 \leq l \leq 1$ . La fonction  $s$  est dite concave si, pour tout  $u, v \in \Omega$ ,  $s(v) \geq (1-l)s(u) + ls(v)$  ([75] page 160).

l'inefficacité relative asymptotique de  $E_n$  par rapport à  $\widehat{F}_{\kappa\Delta}^n$  (I.48) tend vers l'infini quand le nombre d'observations  $n$  augmente avec un choix approprié de noyau, par exemple

$$\kappa(u) = \begin{cases} \frac{9}{8}(1 - \frac{5}{3}u^2) & \text{si } |u| \leq 1, \\ 0 & \text{sinon,} \end{cases} \quad (\text{I.50})$$

avec certaines conditions vérifiées lorsque le support de  $\kappa$  est borné et

$$\int_{\Omega} u\kappa(u)\kappa(u)du > 0. \quad (\text{I.51})$$

Falk a proposé dans [41] une mesure de la performance asymptotique du noyau  $\kappa$  sous la forme :

$$Q(\kappa) = \int_{\Omega} 2\kappa(u)\kappa(u)udu. \quad (\text{I.52})$$

Il est important de noter qu'aucun noyau de carré intégrable ne minimise  $Q$  (I.52) ([101] page 96). En utilisant la mesure de  $R(\kappa) = \int_{\Omega} (\kappa(u))^2 du$  de la performance du noyau en estimation de la densité, Falk a montré dans [41] que le noyau d'Epanechnikov défini par (I.32). Les *noyaux gaussiens* ou *uniformes* ont cependant des performance très proches. Jones a démontré dans [56], en considérant le critère IMSE, que le meilleur noyau est le noyau uniforme. Cependant, les performance d'autre noyaux tels que les noyaux d'Epanechnikov, *normal* et *triangulaire* ne sont que légèrement moins bonnes.

Swanepoel a montré dans [112], en se basant sur le même critère IMSE, que le meilleur noyau est le noyau uniforme défini, pour tout  $v > 0$ , par  $\kappa(u) = \frac{1}{2v}\mathbb{1}_{[-v,v]}(u)$ . Pour les densités de probabilités  $f$  discontinues en un nombre fini de points, le meilleur noyau est le *noyau exponentiel* défini par :  $\kappa(u) = \frac{l}{2}e^{-l|u|}$  pour  $l > 0$ .

Il résulte de ce qui précède que, contrairement à l'estimation par noyau de la densité, le choix optimal du noyau est discutable par rapport au critère d'optimisation utilisé. Loquin et Strauss ont proposé dans [71] une extension de la méthode d'estimation à noyau de la fonction de répartition (I.48) sous le nom d'*estimation imprécise de la fonction de répartition*. Cette extension conduit à une *estimation imprécise de la densité de probabilité* sous-jacente à un ensemble fini d'observations.

# Chapitre II

## Imprécision et mesure de probabilité

L'*imprécision* et l'*incertitude* sont deux aspects de l'imperfection d'une information. Si ces deux aspects peuvent être facilement dissociés dans le langage courant, cette dissociation s'avère plus difficile lorsqu'il s'agit de les représenter à l'aide de la théorie des probabilités. Ce fut une des motivations qui a promu l'émergence de nouvelles théories de représentation d'un défaut de connaissance sur une information.

On appelle incertitude un défaut équivalent à un doute sur la validité d'une connaissance. L'incertitude peut provenir d'un manque de fiabilité de l'intermédiaire d'observation, ce dernier pouvant commettre des erreurs intentionnellement ou non. L'incertitude caractérise aussi la difficulté, pour un utilisateur, à vérifier la pertinence ou la véracité d'une information.

Par opposition, on appelle imprécision un défaut équivalent à une difficulté d'énoncé de la connaissance disponible sur une information. L'imprécision peut venir d'un défaut d'observation (par exemple 15000 à 20000 manifestants), d'une erreur systématique due à l'étalonnage d'un appareil de mesure ( $\pm 0, 1^\circ\text{C}$ ) ou encore à l'énoncé d'une connaissance approximative (prix du carburant dans une région). L'imprécision peut aussi venir d'une connaissance vague énoncée en langage familier (ça coûte cher, c'est loin,...)[36].

Fondamentalement, la *théorie des probabilités* est plutôt dédiée à la gestion des incertitudes tandis que les défauts d'imprécision sont traditionnellement représentés de façon ensembliste (par exemple par l'utilisation d'intervalles de précision) donnant lieu à ce qu'on appelle traditionnellement le calcul d'erreur. Notons qu'il est possible de représenter, via la théorie de probabilités, un défaut d'imprécision par l'intermédiaire des intervalles de confiance [10]. Une dissymétrie fondamentale existe entre ces deux représentations qui est que l'une est nuancée (les probabilités) alors que l'autre est binaire car de nature ensembliste (le calcul d'erreur). L'apparition de la théorie des *sous-ensembles flous*, introduite par Zadeh à partir de 1965 [127], qui propose une généralisation pondérée de la théorie des ensembles, a permis d'introduire des nuances dans le calcul d'erreur par le biais du principe d'extension. La théorie des sous-ensembles flous est un outil mathématique qui définit le concept d'appartenance partielle à un ensemble.

Maintenant, la question qui se pose est : est-il possible de faire cohabiter dans un même formalisme imprécision et incertitude. Dans les premiers modèles proposés, on trouve la proposition de Shafer [102] qui remet en cause le modèle probabiliste en disant que, fondamentalement, il n'y a pas de raison de croire que les jugements d'incertitude devraient être des probabilités. Le cadre probabiliste, d'après Dubois et Prade [35], décrit mal l'ignorance partielle. L'ignorance partielle est justement un défaut s'apparentant plutôt à

l'imprécision.

Au cours des trente dernières années, d'autres théories de l'incertain ont été proposées permettant de prendre en compte l'imprécision au sein de l'incertitude [109]. Parmi ces théories, nous citons les *probabilités imprécises (ou ensembles de probabilités)* [121], la *théorie de fonctions de croyance* [21, 102] et la *théorie de possibilité* [129, 34]. Ces théories, représentent l'incertitude au moyen de deux mesures conjuguées [35, 102]. Les bornes inférieures sont des mesures *sous-additives*, tandis que les bornes supérieures sont des mesures *super-additives* [89].

La théorie des probabilités imprécises proposée par Walley [121] est très certainement le cadre le plus général. Dans cette théorie on suppose qu'il existe une mesure de probabilité précise mais elle n'est pas parfaitement connue [86]. Les manipulations dans ce cadre théorique sont cependant assez complexes. C'est pourquoi des cadres théoriques plus restreints ont été proposés parmi lesquels on retrouve les *p-boxes* [47, 24], les *distributions des probabilités imprécises* [18], . . . . La plupart des autres théories de gestion de l'incertain (et de l'imprécis) peuvent être vues comme des cas particuliers de la théorie des probabilités imprécises. C'est le cas de la théorie des fonctions de croyance [23] introduite par Dempster [21], puis développée par Shafer [102]. La théorie des possibilités, introduite par Zadeh [129] et développée par Dubois et Prade [33] en est aussi un cas particulier, très certainement le plus simple à manipuler mais manquant parfois d'expressivité.

On peut s'interroger sur la nature d'une donnée à la fois imprécise et incertaine. Souvent ces deux notions sont liées car l'incertitude diminue quand l'imprécision augmente. Ainsi, dans un jeu de pile-ou-face, l'événement imprécis  $\{pile, face, tranche\}$  est un événement certain tandis que l'événement  $\{pile, tranche\}$  qui est plus précis est moins certain.

Une modélisation particulièrement pertinente d'une donnée à la fois imprécise et incertaine a émergé à partir des années 1970 sous le nom de *variable aléatoire floue* [16]. Le principe de cette modélisation est d'associer, à chaque expérience aléatoire, un nombre flou en lieu et place d'un nombre précis [76]. Plusieurs visions s'affrontent autour de ce même concept suivant qu'on considère une variable floue comme une généralisation au flou d'une variable aléatoire classique [64, 63], ou comme une généralisation au flou de la théorie des ensembles aléatoires [87].

Reste à mettre en relation le concept de variables aléatoires floues et celui de mesure imprécise de probabilité. Une façon de réaliser cette mise en relation est d'utiliser le principe d'histogramme, permettant de relier la probabilité d'un événement à sa fréquence d'occurrence. Viertl et Trutschnig [120] ont proposé de bâtir une extension cohérente de la notion d'histogramme, aboutissant ainsi à un *histogramme flou*, c'est à dire un histogramme dont les accumulateurs sont des nombres flous. On aboutit alors à définir une *probabilité floue*, qui peut être prise comme un modèle d'ordre deux par rapport aux probabilités imprécises de Walley. Enfin quelques outils sont proposés pour réaliser l'estimation de ces probabilités floues en généralisant au flou des techniques non-paramétriques classiques d'estimation de densité [2].

Rompant plus ou moins avec cette approche, Loquin et Strauss [71] ont montré que l'imprécision des données n'était pas la seule source potentielle d'imprécision des probabilités. Cette imprécision peut aussi provenir du manque de connaissance que l'on a de l'outil le plus adéquat pour estimer la densité de probabilité sous-jacente à un ensemble fini d'observation. Dans ce cas, comme dans les précédents, on aboutit à des modèles de probabilités imprécises [71].

Ce chapitre a pour objet de présenter les propositions de la littérature pour faire cohabiter ces deux aspects d'imprécision et d'incertitude dans le traitement des données réelles. Nous présentons, en section II.1, quelques rappels sur la modélisation assez générale que propose la théorie des sous-ensembles flous pour modéliser l'imprécision en nous focalisant plus particulièrement sur ce que l'on entend par *nombre flou*. D'autre part, nous nous intéressons particulièrement, dans ce chapitre, aux mesures non-additives et à l'extension des opérateurs d'estimation aux probabilités imprécises.

Nous présentons, en section II.2, deux des modèles mathématiques permettant de représenter une variable aléatoire floue, suivant son interprétation. Nous introduisons, dans la même section, le concept de distribution de probabilité floue en s'appuyant, d'une part, sur l'aspect de l'histogramme flou et d'autre part, sur la notion de la *densité de probabilité floue*. Dans cette même section, nous présentons trois extensions de méthodes d'estimation non-paramétrique de la densité considérant un ensemble fini d'observations floues.

Nous présentons, en section II.3, comment une méconnaissance partielle de l'outil le plus adéquat d'estimation de la distribution de probabilité sous-jacente à un ensemble fini d'observations réelles peut être prise en compte pour aboutir à une estimation imprécise de la fonction de répartition. C'est sur ce dernier modèle que s'appuient nos travaux.

## II.1 Quelques rappels

### II.1.1 Sous-ensembles flous

Le concept de sous-ensemble flou<sup>1</sup>, sur lequel s'appuie la logique floue, a été introduit par Zadeh [127] en étendant la théorie des ensembles [55, 79] et la logique propositionnelle classique. La vision la plus connue des sous-ensembles flous est celle de sous-ensembles dont les frontières sont mal connues. Le principe d'extension proposé par Zadeh (énoncé dans [33]) permet de généraliser au flou tout concept mathématique basé sur des sous-ensembles classiques. Les sous-ensembles flous sont généralement utilisés, d'une part, pour leur capacité à modéliser l'imprécision, et d'autre part, parce qu'ils permettent de manipuler mathématiquement des concepts linguistiques.

**Définition II.1 (Sous-ensemble flou)** Soit  $\Omega$  un ensemble de référence. Un "sous-ensemble flou"  $A$  est caractérisé par une "fonction d'appartenance"  $\mu_A$  qui associe, à chaque élément  $x \in \Omega$ , un degré d'appartenance  $\mu_A(x)$  :

$$\mu_A : \begin{array}{l} \Omega \rightarrow [0, 1], \\ x \mapsto \mu_A(x). \end{array} \quad (\text{II.1})$$

Le *noyau* de  $A$  est défini comme l'ensemble des éléments de  $\Omega$  appartenant totalement à  $A$  :

$$\text{noyau}(A) = \{x \in \Omega : \mu_A(x) = 1\}. \quad (\text{II.2})$$

Une  $\alpha$ -coupe de  $A$  (ou coupe de niveau  $\alpha$ ), notée  $A_\alpha$ , est un ensemble ordinaire défini comme l'ensemble des éléments de  $\Omega$  ayant une appartenance supérieure ou égale à  $\alpha$  :

$$\forall \alpha \in (0, 1], A_\alpha = \{x \in \Omega : \mu_A(x) \geq \alpha\}. \quad (\text{II.3})$$

---

1. En anglais *fuzzy set*.

La *hauteur* de  $A$  est la borne supérieure de la fonction d'appartenance  $\mu_A$  :

$$\text{hauteur}(A) = \sup_{x \in \Omega} \mu_A(x). \quad (\text{II.4})$$

Un sous-ensemble flou  $A$  est dit *normalisé* si sa hauteur est égale à 1. Les référentiels que nous considérons plus particulièrement dans ce manuscrit sont des sous-ensembles de  $\mathbb{R}$ . Un sous-ensemble flou d'un ensemble  $\Omega \subseteq \mathbb{R}$  s'appelle une *quantité floue*.

Un *intervalle flou fermé* est défini comme un sous-ensemble flou normalisé de  $\Omega \subseteq \mathbb{R}$  dont les *coupes de niveau* sont des intervalles fermés non-vides. Un *nombre flou* est un cas particulier d'intervalle flou fermé. Plus particulièrement, c'est un intervalle flou dont le noyau se réduit à un singleton [35].

Dans la suite de ce manuscrit, on notera  $\mathcal{H}_c(\Omega)$  la famille des intervalles flous fermés sur  $\Omega$ ,  $\mathcal{K}_c(\Omega)$  la famille des intervalles fermés de  $\Omega$  et  $\mathcal{F}_c(\Omega) \subseteq \mathcal{H}_c(\Omega)$  la famille des nombres flous sur  $\Omega$ .

La coupe de niveau d'un intervalle flou fermé est notée  $A_\alpha = [\underline{a}_\alpha, \bar{a}_\alpha]$  (voir [60]) :

$$\underline{a}_\alpha = \inf \{x \in \Omega : \mu_A(x) \geq \alpha\} \text{ et } \bar{a}_\alpha = \sup \{x \in \Omega : \mu_A(x) \geq \alpha\}. \quad (\text{II.5})$$

Certaines propriétés des intervalles flous sont importantes lorsque cette notion est utilisée à des fins statistiques. Les propriétés suivantes [120, 114] sont des extensions par les coupes de niveau de propriétés des éléments de  $\mathcal{K}_c(\Omega)$  vers les éléments de  $\mathcal{H}_c(\Omega)$ .

Soient  $(a_1, a_2, b_1, b_2) \in \mathbb{R}^4$  tel que  $a_1 \leq a_2$  et  $b_1 \leq b_2$ .  $[a_1, a_2] \in \mathcal{K}_c(\Omega)$  et  $[b_1, b_2] \in \mathcal{K}_c(\Omega)$ . On définit la relation  $\preceq$  par :

$$[a_1, a_2] \preceq [b_1, b_2] :\Leftrightarrow a_1 \leq b_1 \text{ et } a_2 \leq b_2. \quad (\text{II.6})$$

Cette relation induit un ordre partiel sur  $\mathcal{K}_c(\Omega)$  qui peut être étendu à  $\mathcal{H}_c(\Omega)$ , pour tout  $A$  et  $B \in \mathcal{H}_c(\Omega)$ , par :

$$A \preceq B :\Leftrightarrow \forall \alpha \in (0, 1], A_\alpha \preceq B_\alpha. \quad (\text{II.7})$$

L'inclusion dans  $\mathcal{H}_c(\Omega)$  peut être définie, pour tout  $A$  et  $B \in \mathcal{H}_c(\Omega)$ , par :

$$A \subseteq B :\Leftrightarrow \forall \alpha \in (0, 1], A_\alpha \subseteq B_\alpha. \quad (\text{II.8})$$

De même les notions de somme  $\oplus$  (rsp. différence  $\ominus$ ) de Minkowski<sup>2</sup> peuvent être facilement étendues aux intervalles flous via les coupes de niveau, c'est-à-dire :

$$\forall \alpha \in (0, 1], (A \oplus B)_\alpha = A_\alpha \oplus B_\alpha \text{ et } (A \ominus B)_\alpha = A_\alpha \ominus B_\alpha. \quad (\text{II.9})$$

La distance entre deux intervalles flous fermés  $A$  et  $B$  de  $\mathcal{H}_c(\Omega)$  est définie par :

$$d_{H,\infty}(A, B) := \sup_{\alpha \in (0, 1]} d_H(A_\alpha, B_\alpha), \quad (\text{II.10})$$

où  $d_H$  est la distance de Hausdorff [97] définie, pour tout  $A, B \in \mathcal{K}_c(\Omega)$ , par :

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} |a - b|, \sup_{b \in B} \inf_{a \in A} |a - b| \right\}. \quad (\text{II.11})$$

---

2. La somme de Minkowski de deux intervalles  $A$  et  $B$  est donnée par :  $A \oplus B = [a_1, a_2] \oplus [b_1, b_2] = [a_1 + b_1, a_2 + b_2]$ . La différence de Minkowski de  $A$  et  $B$  est définie, à partir de la somme de Minkowski, par :  $A \ominus B = A \oplus (-B)$  [92].

## II.1.2 Mesures de confiances non-additives

### II.1.2.1 Capacité

La notion de *capacité* a été introduite par Choquet [13] dans le contexte de sa théorie des capacités. Un concept similaire a été proposé par Sugeno [111] sous le nom de *mesure floue* et par Denneberg [22] sous le nom de *mesure non-additive*.

**Définition II.2 (Capacité [13, 102])** Soit  $\mathcal{P}(\Omega)$  l'ensemble des parties de  $\Omega$ , la fonction  $\nu : \mathcal{P}(\Omega) \rightarrow [0, 1]$  est une capacité sur  $\Omega$ , si  $\nu$  vérifie :

1.  $\nu$  est normalisée, c'est-à-dire  $\nu(\emptyset) = 0$  et  $\nu(\Omega) = 1$ .
2.  $\nu$  est monotone, c'est-à-dire  $\forall A, B \in \mathcal{P}(\Omega), A \subseteq B \Rightarrow \nu(A) \leq \nu(B)$ .

Une capacité  $\nu$  est dite *k-monotone*, avec  $k \in \mathbb{N}$  et  $k \geq 2$ , si pour toute famille  $(A_i)_{i=1, \dots, k}$  de  $\Omega$  on a :

$$\nu\left(\bigcup_{i=1}^k A_i\right) + \sum_{I \subseteq \{1, \dots, k\}: I \neq \emptyset} (-1)^{|I|} \nu\left(\bigcap_{i \in I} A_i\right) \geq 0, \quad (\text{II.12})$$

où  $|\cdot|$  est le cardinal d'un ensemble.

Cette capacité est dite  $\infty$ -monotone si elle est monotone pour tout  $k$ . Une capacité  $\nu$  2-monotone est dite *convexe* ou *super-additive*. Elle vérifie par ailleurs :

$$\forall A, B \in \mathcal{P}(\Omega), \nu(A \cup B) + \nu(A \cap B) \geq \nu(A) + \nu(B). \quad (\text{II.13})$$

Une capacité  $\nu$  est dite *concave* ou *sous-additive*, si :

$$\forall A, B \in \mathcal{P}(\Omega), \nu(A \cup B) + \nu(A \cap B) \leq \nu(A) + \nu(B). \quad (\text{II.14})$$

La *capacité conjuguée*  $\nu^c$  d'une capacité  $\nu$  est définie par :

$$\forall A \in \mathcal{P}(\Omega), \nu^c(A) = 1 - \nu(A^c). \quad (\text{II.15})$$

où  $A^c$  est le complémentaire de  $A$  dans  $\Omega$ .

Remarquons que si  $\nu$  est une capacité concave,  $\nu^c$  est une capacité convexe et vice-versa.

Le *coeur*<sup>3</sup> d'une capacité concave  $\nu$ , que nous notons  $\mathcal{M}(\nu)$ , est la famille des mesures de probabilités définie par :

$$\mathcal{M}(\nu) = \left\{ P \in \mathbb{P}(\Omega) : \forall A \subseteq \Omega, P(A) \leq \nu(A) \right\}, \quad (\text{II.16})$$

où  $\mathbb{P}(\Omega)$  est l'ensemble des mesures de probabilité définies sur  $\Omega$ .

La famille  $\mathcal{M}(\nu)$  peut aussi s'écrire :

$$\mathcal{M}(\nu) = \left\{ P \in \mathbb{P}(\Omega) : \forall A \subseteq \Omega, \nu^c(A) \leq P(A) \leq \nu(A) \right\}, \quad (\text{II.17})$$

où  $\nu^c(A) = 1 - \nu(A^c)$ .

On peut remarquer facilement qu'une mesure de probabilité est un cas particulier de capacité. Son coeur se réduit à un élément : elle-même. Une mesure de probabilité est une capacité qui vérifie l'axiome d'additivité :

$$\forall A, B \subseteq \Omega, A \cap B = \emptyset, P(A \cup B) = P(A) + P(B). \quad (\text{II.18})$$

Un corollaire de cette propriété est que la capacité complémentaire d'une mesure de probabilité est elle-même. Une probabilité est une capacité à la fois convexe et concave.

---

3. En anglais *core*.



### II.1.2.2 Intégrale de Choquet et espérance imprécise

L'intégrale de Choquet permet d'étendre, aux mesures non-additives, le concept d'intégrale par rapport à une mesure [13, 22]. Elle permet ainsi de définir l'espérance mathématique (cf. section I.1) d'une variable aléatoire vis à vis d'une capacité. L'intégrale de Choquet peut être vue comme une généralisation de l'opérateur d'espérance mathématique [80].

**Définition II.3 (Intégrale de Choquet continue [13, 11])** Soit  $\nu$  une capacité sur  $\Omega$  et soit  $s$  une fonction bornée de  $L(\Omega)$ , l'intégrale de Choquet continue de  $s$  par rapport à  $\nu$  est définie par :

$$(C) \int_{\Omega} s d\nu = \int_{-\infty}^0 \left( \nu(\{u \in \Omega : s(u) \geq \alpha\}) - 1 \right) d\alpha + \int_0^{\infty} \nu(\{u \in \Omega : s(u) \geq \alpha\}) d\alpha. \quad (\text{II.19})$$

On peut trouver dans ([69] page 66) le passage de la forme continue à la forme discrète pour une intégrale de Choquet :

**Définition II.4 (Intégrale de Choquet discrète [13, 11])** Soit  $\nu$  une capacité sur  $\Omega$  et soit  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ , l'intégrale de Choquet discrète de  $x$  par rapport à  $\nu$  est définie par :

$$(C) \int_{\Omega} x d\nu = \sum_{i=1}^n (x_{\tau(i)} - x_{\tau(i-1)}) \nu(A_i), \quad (\text{II.20})$$

où  $\tau$  est la permutation sur  $\Omega$ , telle que  $x_{\tau(1)} \leq \dots \leq x_{\tau(n)}$  avec, par convention,  $x_{\tau(0)} = 0$  et  $A_i := \{\tau(i), \dots, \tau(n)\}$ .

Dans la suite de ce manuscrit, pour une fonction bornée  $s$  de  $L(\Omega)$  et une capacité  $\nu$  de  $\Omega$ , on notera  $\mathbb{C}_{\nu}(s)$  l'intégrale de Choquet de  $s$  par rapport à  $\nu$ .

On peut remarquer facilement que, si la capacité  $\nu$  est une mesure de probabilité  $P$  alors l'intégrale de Choquet de la fonction bornée  $s$  de  $L(\Omega)$  devient simplement l'espérance mathématique (Définition I.6) de  $s$  dans un voisinage probabiliste défini par  $P$ . En d'autres termes, pour les mesures additives, l'intégrale de Choquet coïncide avec l'intégrale de Lebesgue.

Un résultat essentiel pour les capacités, connu sous le nom de *Théorème de Schmeidler-Denneberg*, a permis Rico et Strauss [92] de définir une extension de l'opérateur espérance mathématique sous le nom d'*espérance imprécise*. Ce théorème est dû à Schmeidler ([96] proposition 3) et Denneberg ([22] proposition 10.3).

**Théorème II.5 (Théorème de Schmeidler-Denneberg)** Soient  $\nu$  une capacité concave sur  $\Omega$  et  $\mathcal{M}(\nu)$  la famille des mesures de probabilités définie par (II.26) et soit  $s$  une fonction bornée de  $L(\Omega)$ , alors :

$$\mathbb{C}_{\nu^c}(s) = \inf_{P \in \mathcal{M}(\nu)} \mathbb{E}_P(s), \quad (\text{II.21})$$

et

$$\mathbb{C}_{\nu}(s) = \sup_{P \in \mathcal{M}(\nu)} \mathbb{E}_P(s). \quad (\text{II.22})$$

où  $\mathbb{E}_P(\cdot)$  est l'espérance mathématique par rapport à la mesure de probabilité  $P$ .

L'espérance imprécise définie par Rico et Strauss [92] utilise le Théorème II.5 :

**Définition II.6 (Espérance imprécise au sens de Rico-Strauss)** Soit  $\nu$  une capacité concave sur  $\Omega$  et soit  $s$  une fonction bornée de  $L(\Omega)$ , l'espérance imprécise de  $s$  basée sur  $\nu$ , notée  $\underline{\mathbb{E}}_\nu(s)$ , est définie par :

$$\underline{\mathbb{E}}_\nu(s) = [\underline{\mathbb{E}}_\nu(s), \overline{\mathbb{E}}_\nu(s)] = \{\mathbb{E}_P(s) : P \in \mathcal{M}(\nu)\}, \quad (\text{II.23})$$

où  $\underline{\mathbb{E}}_\nu(s) = \mathbb{C}_{\nu^c}(s)$  (resp.  $\overline{\mathbb{E}}_\nu(s) = \mathbb{C}_\nu(s)$ ) est la borne inférieure (resp. supérieure) de  $\underline{\mathbb{E}}_\nu(s)$ .

De cette définition découle la propriété de domination suivante :

$$\forall y \in \underline{\mathbb{E}}_\nu(s), \exists P \in \mathcal{M}(\nu) : y = \mathbb{E}_P(s). \quad (\text{II.24})$$

### II.1.2.3 Ensembles de probabilités

Lorsque la probabilité sous-jacente à un ensemble fini d'observations n'est pas parfaitement connue, une des manières de représenter cette méconnaissance est d'utiliser des ensembles de probabilités à la place de probabilités précises [105]. Ce modèle, dû à Walley [121], qui propose une représentation plus générale de l'incertitude par une paire duale de mesures de confiances : les *probabilités inférieures et supérieures*.

La probabilité inférieure  $\underline{P}$  est définie ici comme une capacité convexe (II.13). Tandis que la probabilité supérieure, notée  $\overline{P}$ , est définie comme la capacité concave (II.14) (conjuguée de la capacité convexe  $\underline{P}$ ). La probabilité inférieure est liée à la probabilité supérieure par :

$$\forall A \subseteq \Omega, \overline{P}(A) = 1 - \underline{P}(A^c). \quad (\text{II.25})$$

On peut définir à partir de  $\underline{P}$  et  $\overline{P}$  une famille de mesures de probabilité, notée  $\mathcal{M}(\overline{P})$ , par :

$$\begin{aligned} \mathcal{M}(\overline{P}) &= \left\{ P \in \mathbb{P}(\Omega) : \forall A \subseteq \Omega, P(A) \leq \overline{P}(A) \right\} \\ &= \left\{ P \in \mathbb{P}(\Omega) : \forall A \subseteq \Omega, \underline{P}(A) \leq P(A) \leq \overline{P}(A) \right\}, \end{aligned} \quad (\text{II.26})$$

avec  $\underline{P}(A) = 1 - \overline{P}(A^c)$ .

Les probabilités supérieures (resp. inférieures) considérées ici sont les enveloppes de l'ensemble de probabilités induit (probabilités supérieures cohérentes) (resp. probabilités inférieures cohérentes) [24].

Nous présentons dans la suite quatre modèles moins généraux avec l'intention de fournir des outils pratiques de manipulation d'incertitude : les *p-box*, les *distributions imprécises de probabilités*, les *fonctions de croyance* et les *distributions de possibilité*.

#### II.1.2.3.1 P-boxes

Une p-box est une paire de fonctions  $(\underline{F}, \overline{F})$ , telles que  $\forall x \in \Omega, \underline{F}(x) \leq \overline{F}(x)$ . Une p-box [47, 42] peut être vue comme une classe de mesures de probabilité dont les fonctions de répartitions sont bornées par  $\underline{F}$  et  $\overline{F}$ . La famille de probabilité définie à partir d'une p-box  $(\underline{F}, \overline{F})$  peut alors être notée  $\mathcal{M}(\underline{F}, \overline{F})$  et donnée par :

$$\mathcal{M}(\underline{F}, \overline{F}) = \left\{ P \in \mathbb{P}(\Omega) : \forall x \in \Omega, \underline{F}(x) \leq P((-\infty, x]) \leq \overline{F}(x) \right\}. \quad (\text{II.27})$$

Les p-boxes sont définies à l'origine sur des ensembles réels. On doit à Destercke [24] leur généralisation à n'importe quel ensemble. Ce modèle est simple à manipuler et permet de représenter une méconnaissance de la borne de distribution de probabilité modélisant un problème [109]. Notons que l'écart entre les bornes  $\underline{F}$  et  $\overline{F}$  exprime le caractère imprécis de l'information [6].

### II.1.2.3.2 Distributions imprécises de probabilité

Une distribution imprécise de probabilité [18] est définie en tout point par deux fonctions notées  $\mathfrak{l}$  (pour lower) et  $\mathfrak{u}$  (pour upper) telles que  $\forall x \in \Omega, \mathfrak{l}(x) \leq \mathfrak{u}(x)$ . On peut définir, à partir de la paire de fonctions  $(\mathfrak{l}, \mathfrak{u})$ , une famille de probabilité notée  $\mathcal{M}((\mathfrak{l}, \mathfrak{u}))$ , par :

$$\mathcal{M}((\mathfrak{l}, \mathfrak{u})) = \left\{ P \in \mathbb{P}(\Omega) : \forall x \in \Omega, \mathfrak{l}(x) \leq f_P(x) \leq \mathfrak{u}(x) \right\}, \quad (\text{II.28})$$

où  $f_P$  est la densité de probabilité associée à la mesure de probabilité  $P$ .

Il est important de noter que, contrairement aux p-boxes qui sont définies par deux fonctions de répartitions,  $\mathfrak{l}$  et  $\mathfrak{u}$  ne sont pas des densités de probabilités. Le calcul de la probabilité imprécise associée à un ensemble  $A \subseteq \Omega$  est obtenu par :

$$\underline{P}(A) = \max \left( \int_A \mathfrak{l}(x) dx, 1 - \int_{A^c} \mathfrak{u}(x) dx \right), \quad (\text{II.29})$$

et

$$\overline{P}(A) = \min \left( \int_A \mathfrak{u}(x) dx, 1 - \int_{A^c} \mathfrak{l}(x) dx \right). \quad (\text{II.30})$$

### II.1.2.4 Théorie des fonctions de croyance

Un *sous-ensemble aléatoire* est une fonction de masse  $m : \mathcal{P}(\Omega) \rightarrow [0, 1]$  des sous-ensembles de  $\Omega$  à valeurs dans l'intervalle  $[0, 1]$  telle que  $\sum_{A \subseteq \Omega} m(A) = 1$ ,  $m(A) \geq 0$  et  $m(\emptyset) = 0$  [25]. La masse  $m(A)$  s'interprète comme la probabilité de savoir uniquement que la vraie valeur se trouve dans  $A$ . Un sous-ensemble  $A$  tels que  $m(A) > 0$  est appelé *ensemble focal*. A partir de cette fonction, Shafer a défini dans [102] deux fonctions d'ensembles, la mesure de *croyance* et la mesure de *plausibilité*. Ces deux mesures sont données [102] :

$$\forall A \subseteq \Omega, \text{Bel}(A) = \sum_{B, B \subseteq A} m(B) \quad (\text{croyance}), \quad (\text{II.31})$$

et

$$\text{Pl}(A) = \sum_{B, B \cap A \neq \emptyset} m(B) \quad (\text{plausibilité}). \quad (\text{II.32})$$

La mesure de plausibilité est liée à la mesure de croyance par :

$$\forall A \subseteq \Omega, \text{Pl}(A) = 1 - \text{Bel}(A^c). \quad (\text{II.33})$$

La quantité  $\text{Bel}(A)$  représente la croyance totale en  $A \subseteq \Omega$ . La fonction de croyance  $\text{Bel}$  vérifie la propriété de super-additivité (II.13) [86].

La quantité  $\text{Pl}(A)$  représente la croyance qui pourrait être attribuée à  $A$ , compte tenu des éléments qui ne contredisent pas cette proposition. La fonction de plausibilité  $\text{Pl}(A)$  est une capacité sous-additive (II.14) [86].

On peut définir à partir, de  $Pl$  et  $Bel$ , une famille de probabilité, notée  $\mathcal{M}(Pl)$ , par :

$$\begin{aligned}\mathcal{M}(Pl) &= \left\{ P \in \mathbb{P}(\Omega) : \forall A \subseteq \Omega, P(A) \leq Pl(A) \right\} \\ &= \left\{ P \in \mathbb{P}(\Omega) : \forall A \subseteq \Omega, Bel(A) \leq P(A) \leq Pl(A) \right\}.\end{aligned}\quad (\text{II.34})$$

Notons que, pour un sous-ensemble  $A \subseteq \Omega$ , l'écart entre les deux bornes  $Pl(A)$  et  $Bel(A)$  de l'intervalle  $[Bel(A), Pl(A)]$  mesure l'ignorance relative à  $A$ .

Les fonctions de croyance peuvent être vues comme des distributions de probabilités portant sur des ensembles [25]. Les techniques de manipulations des fonctions de croyance via leur représentations sur les masses sont assez simples mais conduisent à des algorithmiques assez lourdes [109].

### II.1.2.5 Théorie de possibilités

La théorie des possibilités, introduite par Zadeh [129], est basée sur la théorie des sous-ensembles flous [34, 128]. Une *distribution de possibilités*, qui induit une mesure de possibilité, est un modèle simple permettant de modéliser des informations imprécises [24]. On doit à Dubois et Prade [33] de nombreux développements autour de cette notion de distribution de possibilité. Certains de leurs travaux montrent que les mesures de possibilités permettent de représenter facilement des notions statistiques complexes [109]. Dubois a montré, dans [32], que les intervalles de confiance [10] issus d'une distribution de probabilité ne sont autre qu'une distribution de possibilité.

Une mesure de possibilité est une capacité  $\Pi : \mathcal{P}(\Omega) \rightarrow [0, 1]$ , vérifiant :

$$\forall A, B \subseteq \Omega, \Pi(A \cup B) = \max(\Pi(A), \Pi(B)). \quad (\text{II.35})$$

Le nombre  $\Pi(A)$  quantifie dans quelle mesure l'événement  $A$  est possible dans  $\Omega$ . Une mesure de possibilité  $\Pi$  peut être caractérisée par une distribution de possibilité  $\pi : \Omega \rightarrow [0, 1]$  telle que  $\sup_{x \in \Omega} \pi(x) = 1$  et qui vérifie :

$$\forall A \subseteq \Omega, \Pi(A) = \sup_{x \in A} \pi(x). \quad (\text{II.36})$$

Une distribution de possibilité est équivalente à un sous-ensemble flou normalisé. Elle peut aussi s'interpréter comme un ensemble d'intervalles de confiances emboîtés [37]. Cette distribution de possibilité définit une mesure de confiance duale à la mesure de possibilité, appelée mesure de nécessité :

$$\forall A \subseteq \Omega, N(A) = 1 - \Pi(A^c). \quad (\text{II.37})$$

La mesure de nécessité est calculée, via la distribution de possibilité  $\pi$ , par :

$$\forall A \subseteq \Omega, N(A) = \inf_{x \notin A} (1 - \pi(x)). \quad (\text{II.38})$$

La mesure de nécessité vérifie :

$$\forall A, B \subseteq \Omega, N(A \cap B) = \min(N(A), N(B)). \quad (\text{II.39})$$

Une mesure de possibilité  $\Pi$  (rsp. nécessité  $N$ ) est une capacité concave (II.14) (rsp. convexe (II.13)). On peut définir, à partir de  $\Pi$  et  $N$ , une famille de mesures de probabilité, notée  $\mathcal{M}(\Pi)$ , par :

$$\begin{aligned}\mathcal{M}(\Pi) &= \left\{ P \in \mathbb{P}(\Omega) : \forall A \subseteq \Omega, P(A) \leq \Pi(A) \right\} \\ &= \left\{ P \in \mathbb{P}(\Omega) : \forall A \subseteq \Omega, N(A) \leq P(A) \leq \Pi(A) \right\}.\end{aligned}\quad (\text{II.40})$$

Notons que l'écart entre les deux bornes  $\Pi$  et  $N$  de l'intervalle  $[\Pi, N]$  exprime le caractère imprécis de l'information [6].

Par abus de notation, on notera  $\mathcal{M}(\pi)$  la famille de probabilité définie par (II.40), où  $\pi$  est la distribution de possibilité induite par la mesure de possibilité  $\Pi$ .

## II.2 Mesure de probabilité imprécise et données imprécises

Dans de nombreuses expériences les observations sont des valeurs réelles supposées précises (taille, poids, ...). Lorsque l'issue d'une expérimentation n'est pas une valeur réelle, son traitement statistique est rendu plus aisé en associant, à chaque issue de l'expérimentation, une valeur réelle. Cette association est appelée *variable aléatoire (classique)*. Une variable aléatoire précise est un nombre réel associé au résultat d'une épreuve.

Comme le fait remarquer Kwakernaak [64], lorsque les observations sont imprécises une façon de prendre en compte cette imprécision est d'associer, à chaque résultat d'une expérience aléatoire, un nombre flou plutôt qu'un nombre précis. Cette association porte alors le nom de variable aléatoire floue (VAF)<sup>4</sup>. En d'autres termes, une variable aléatoire floue, d'après Kwakernaak [64], est une généralisation au flou d'une variable aléatoire classique [103, 16]. Dans la même optique, Puri et Ralescu [87], voient une variable aléatoire floue comme une généralisation au flou d'un ensemble aléatoire [103, 49].

Une variable aléatoire floue, d'après Viertl et Trutschnig [120], induit une *distribution de probabilité floue*. Trutschnig a proposé dans [114] d'estimer cette probabilité floue par un histogramme flou. Ce dernier, d'après Trutschnig [114], est une extension au flou de l'histogramme classique en s'appuyant sur les coupes de niveau. Dans la même optique, Viertl [118, 117] définit une densité de probabilité floue à partir d'une procédure d'*intégrale floue*. Cette densité de probabilité floue induit, d'après Viertl [117], une distribution de probabilité floue.

Dans ce contexte, Arefi et al. ont proposé dans [2] une extension au flou de trois méthodes d'estimation non-paramétrique classiques, présentées en section I.2, d'une densité de probabilité sous-jacente à un ensemble fini d'observations floues. Cette extension se base sur l'utilisation d'une décomposition par coupe de niveau.

Nous présentons, dans cette section, deux visions du même concept de variable aléatoire floue : la première est proposée par Puri et Ralescu [87] et la deuxième est issue des travaux de Kwakernaak [64] puis Kruse et Meyer [63]. Nous présentons ensuite le concept d'un histogramme flou au sens de Viertl et Trutschnig [120] ainsi que le concept de *distribution de probabilité floue*. En ce sens, un histogramme flou, d'après Trutschnig

---

4. En anglais Fuzzy Random Variable (FRV).

[114], peut être vu comme un estimateur d'une distribution de probabilité floue quand le nombre d'observations floues tend vers l'infini. Nous présentons, à la fin de cette section, l'extension au flou de trois méthodes d'estimation non-paramétrique (classique) proposées par Arefi et al. [2].

### II.2.1 Variables aléatoires floues

Une variable aléatoire floue peut être vue comme la généralisation au flou d'une variable aléatoire classique [76, 103] ou d'un ensemble aléatoire [103]. Cette généralisation consiste à associer, à toute issue possible d'une expérience aléatoire, un nombre flou en lieu et place d'un nombre réel. On peut retrouver plusieurs cadres de définition pour une variable aléatoire floue. Au sens de Puri et Ralescu [87], une variable aléatoire floue est un ensemble emboîté de variables aléatoires. En ce sens, une variable aléatoire floue peut être vue comme un mécanisme ayant pour résultat expérimental un sous-ensemble flou de  $\mathcal{F}_c(\Omega)$  [49, 15], c'est-à-dire un nombre flou.

Rompant avec cette vision, Kwakernaak [64, 65] propose de voir une variable aléatoire floue comme la perception floue d'une variable aléatoire précise [45, 46]. Ce concept a été formalisé par Kruse et Meyer [63] aboutissant à un modèle mathématique mettant l'accent sur les bornes de la variable aléatoire [46].

**Définition II.7 (VAF au sens de Puri-Ralescu)** Soit  $(\Omega, \mathcal{A})$  un espace mesurable, l'application  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  est une variable aléatoire floue associée à  $(\Omega, \mathcal{A})$  si, pour tout  $\alpha \in (0, 1]$ , l'application

$$\mathcal{X}_\alpha : \begin{array}{l} \Omega \rightarrow \mathcal{K}_c(\Omega) \\ x \mapsto \mathcal{X}_\alpha(x) = (\mathcal{X}(x))_\alpha, \end{array} \quad (\text{II.41})$$

est un ensemble aléatoire (Définition I.4),  $\mathcal{X}_\alpha$  étant la coupe de niveau  $\alpha$  de  $\mathcal{X}$  (II.5) [45].

En se basant sur cette définition, Trutschnig a proposé dans [114] la reformulation suivante :

**Définition II.8.** Soit  $(\Omega, \mathcal{A})$  un espace mesurable et soit  $\mathcal{B}(\Omega)$  une tribu borélienne sur  $\Omega$ , l'application  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  est une variable aléatoire floue si  $\forall A \in \mathcal{B}(\Omega)$  on a :

$$\{x \in \Omega : \forall \alpha \in (0, 1], \mathcal{X}_\alpha(x) \cap A \neq \emptyset\} \in \mathcal{A}. \quad (\text{II.42})$$

**Définition II.9 (VAF au sens de Kwakernaak et Kruse-Meyer)** L'application  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  est une variable aléatoire floue si, pour tout  $\alpha \in (0, 1]$ , les applications

$$\inf \mathcal{X}_\alpha : \begin{array}{l} \Omega \rightarrow \mathbb{R} \\ x \mapsto \inf \mathcal{X}_\alpha(x) = \inf (\mathcal{X}(x))_\alpha, \end{array} \quad (\text{II.43})$$

et

$$\sup \mathcal{X}_\alpha : \begin{array}{l} \Omega \rightarrow \mathbb{R} \\ x \mapsto \sup \mathcal{X}_\alpha(x) = \sup (\mathcal{X}(x))_\alpha, \end{array} \quad (\text{II.44})$$

sont deux variables aléatoires réelles, avec :

$$\mathcal{X}_\alpha(x) = [\inf (\mathcal{X}(x))_\alpha, \sup (\mathcal{X}(x))_\alpha].$$

Dans la suite de ce chapitre, nous considérons les variables aléatoires floues définies au sens de Puri-Ralescu (Définition II.7). Soit  $\mathcal{X}$  une variable aléatoire floue, les coupes de niveau  $\alpha$  de  $\mathcal{X}$  seront notées  $\mathcal{X}_\alpha = [\underline{x}_\alpha, \bar{x}_\alpha]$ . Les terminologies suivantes des variables aléatoires floues sont issues de [2, 114].

**Définition II.10 (VAF indépendantes [2])** Soient  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  et  $\mathcal{Y} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  deux variables aléatoires floues, on dit que  $\mathcal{X}$  et  $\mathcal{Y}$  sont indépendantes si et seulement si chaque variable aléatoire réelle de l'ensemble  $\{\underline{x}_\alpha, \bar{x}_\alpha, \forall \alpha \in (0, 1]\}$  est indépendante de chaque variable aléatoire réelle de l'ensemble  $\{\underline{y}_\alpha, \bar{y}_\alpha, \forall \alpha \in (0, 1]\}$ .

Trutschnig voit dans [114] l'indépendance de deux variables aléatoires floues en terme probabiliste :

**Définition II.11 (VAF indépendantes [114])** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et soit  $\mathcal{B}(\Omega)$  une tribu borélienne sur  $\Omega$ . Soient  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  et  $\mathcal{Y} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  deux variables aléatoires floues, on dit que  $\mathcal{X}$  et  $\mathcal{Y}$  sont indépendantes si, pour tout  $\alpha \in (0, 1]$  et  $A, B \in \mathcal{B}(\Omega)$ , on a :

$$\forall x \in \Omega, P(\mathcal{X}_\alpha(x) \subseteq A, \mathcal{Y}_\alpha(x) \subseteq B) = P(\mathcal{X}_\alpha(x) \subseteq A)P(\mathcal{Y}_\alpha(x) \subseteq B). \quad (\text{II.45})$$

**Définition II.12 (VAF identiquement distribuées [2])** Soient  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  et  $\mathcal{Y} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  sont deux variables aléatoires floues, on dit que  $\mathcal{X}$  et  $\mathcal{Y}$  sont identiquement distribuées si l'ensemble  $\{\underline{x}_\alpha, \bar{x}_\alpha, \forall \alpha \in (0, 1]\}$  et l'ensemble  $\{\underline{y}_\alpha, \bar{y}_\alpha, \forall \alpha \in (0, 1]\}$  sont identiquement distribuées.

Dans la même optique, Viertl [120, 117] et Trutschnig [113, 114] ont proposé de représenter, par une densité de probabilité dont la valeur serait floue (*densité de probabilité floue*), la mauvaise connaissance, due au défaut de mesure, de la densité de probabilité associée à une variable aléatoire précise. Dans son approche, Viertl [117] a montré que les distributions de probabilités floues peuvent être induites par des variables aléatoires floues  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Trutschnig a montré dans [114] que cette distribution de probabilité floue est la limite d'un histogramme flou.

## II.2.2 Histogramme flou et distribution de probabilité floue

Pour appréhender de façon simple la distribution de probabilité d'un ensemble fini d'observations  $(x_1, \dots, x_n)$  de  $n$  variables aléatoires  $(X_1, \dots, X_n)$  i.i.d sur  $\Omega$ , une des méthodes les plus classiques consiste à en réaliser un histogramme (cf. section I.2.1.1). Réaliser un histogramme d'observations réelles consiste à subdiviser un sous-ensemble  $\Omega \subseteq \mathbb{R}$  en  $p \in \mathbb{N}$  sous-ensembles disjoints formant une partition (c'est-à-dire que l'union des  $p$  sous-ensemble est égale à  $\Omega$  et l'intersection de deux sous-ensembles disjoints de la partition est vide). Une propriété utile des histogrammes, les reliant aux distributions de probabilité, est qu'une distribution de probabilité peut être vue comme la limite de l'histogramme (cf. section I.2.1.2) quand le nombre d'observations tend vers l'infini et la granularité de la partition tend vers zéro.

Le principe des histogrammes classiques repose sur le comptage, pour chaque cellule de la partition considérée, du nombre d'observation lui appartenant. Lorsque les observations ne sont pas ponctuelles mais floues, rares sont les observations complètement incluses dans

une des cellules de la partition. Les histogrammes classiques, dans ce cas, ne peuvent pas représenter fidèlement ce type de données. Il est donc nécessaire d'étendre, de façon rigoureuse, la notion d'histogramme pour prendre en compte le caractère flou des données.

Niculescu et Viertl [83], d'une part, et Viertl et Trutschnig [120], d'autre part, ont proposé, sous le nom de histogramme flou, une telle extension. Comme le font remarquer Viertl et Trutschnig [120], la notion de la distribution de probabilité peut aussi être étendue sous le nom de distribution de probabilité floue.

Dans cette section, nous présentons l'histogramme flou au sens de Viertl-Trutschnig [120] ainsi que ses propriétés. Nous présentons également, la notion de la distribution de probabilité floue au sens de Trutschnig [114] et Viertl [117].

### II.2.2.1 Histogrammes flous

Soit  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  un ensemble fini d'observations floues (nombres flous) de  $n$  variables aléatoires floues  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$ . La construction d'un histogramme flou, au sens de Viertl-Trutschnig [120], s'appuie sur de notions de *fréquence relative floue*. Soit  $A \subseteq \Omega$ , la fréquence relative floue de  $A$  est un nombre flou  $\mathcal{H}^n(A) \in \mathcal{F}_c(\Omega)$  défini par ses coupes de niveau par :

$$\forall \alpha \in (0, 1], \mathcal{H}_\alpha^n(A) = \left[ \underline{h}_\alpha^n(A), \overline{h}_\alpha^n(A) \right], \quad (\text{II.46})$$

où  $\underline{h}_\alpha^n(A)$ , la fréquence relative inférieure de niveau  $\alpha$  de l'ensemble  $A \subseteq \Omega$ , est le rapport du nombre de données dont les coupes de niveau  $\alpha$  sont incluses dans  $A$  sur le nombre total de données :

$$\forall \alpha \in (0, 1], \underline{h}_\alpha^n(A) = \frac{1}{n} \left| \{i : \mathcal{X}_{i\alpha} \subseteq A\} \right|, \quad (\text{II.47})$$

et  $\overline{h}_\alpha^n(A)$ , la fréquence relative supérieure de niveau  $\alpha$  de l'ensemble  $A$ , est le rapport du nombre de données dont la coupe de niveau  $\alpha$  a une intersection non vide avec  $A$  sur le nombre total de données :

$$\forall \alpha \in (0, 1], \overline{h}_\alpha^n(A) = \frac{1}{n} \left| \{i : \mathcal{X}_{i\alpha} \cap A \neq \emptyset\} \right|. \quad (\text{II.48})$$

On peut remarquer facilement que, pour tout  $n$  et  $A$  fixés,  $\underline{h}_\alpha^n(A)$  est une fonction monotone décroissante tandis que  $\overline{h}_\alpha^n(A)$  est une fonction monotone croissante en  $\alpha$  [114]. Par conséquence :

$$\forall \alpha, \beta \in (0, 1], \text{ si } \alpha \leq \beta, \text{ alors } \left[ \underline{h}_\beta^n(A), \overline{h}_\beta^n(A) \right] \subseteq \left[ \underline{h}_\alpha^n(A), \overline{h}_\alpha^n(A) \right]. \quad (\text{II.49})$$

Un certain nombre de propriétés de la fréquence relative floue (II.46) issues de [114] sont données :

**Propriété II.13 (Propriétés de la fréquence relative floue)** Soit  $\mathcal{H}^n(\cdot)$  la fréquence relative floue définie par (II.46) et soient  $A, B \subseteq \Omega$ . On a :

1.  $\text{supp}(\mathcal{H}^n(A)) \subseteq [0, 1]$ , où  $\text{supp}(\cdot)$  est le support<sup>5</sup> de fonction.

---

5. Soit  $\varphi \in L(\Omega)$ , on appelle support d'une fonction  $\varphi$ , noté  $\text{supp}(\varphi)$ , le plus petit fermé contenant toutes les valeurs de  $u \in \Omega$  telles que  $\varphi(u) \neq 0$  ([54] page 211).



2.  $\mathcal{H}^n(\Omega) = \mathbb{1}_{\{1\}}$ ,  $\mathcal{H}^n(\emptyset) = \mathbb{1}_{\{0\}}$ , où  $\mathbb{1}_A(\cdot)$  est la fonction indicatrice de  $A$  (I.3).
3. Si  $A \subseteq B \subseteq \Omega$ , alors  $\mathcal{H}^n(A) \preceq \mathcal{H}^n(B)$ , où  $\preceq$  est la relation définie par (II.7).
4. Si  $A \cap B = \emptyset$ , alors  $\mathcal{H}^n(A \cup B) \subseteq \mathcal{H}^n(A) \oplus \mathcal{H}^n(B)$ , où  $\oplus$  est la somme de Minkowski (II.9).
5.  $\mathcal{H}^n(A^c) = \mathbb{1}_{\{1\}} \ominus \mathcal{H}^n(A)$ , où  $A^c$  est le complémentaire de  $A$  de  $\Omega$  et  $\ominus$  est la différence de Minkowski (II.9).

Il est important de noter qu'il existe d'autres définitions de la fréquence relative floue. Par exemple, au sens de Niculescu-Viertl [83], la fréquence relative floue d'un ensemble  $A \subseteq \Omega$ , notée  $\mathcal{H}_A$ , considérant un ensemble fini d'observations floues  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$ , est une fonction à valeurs réelles sur l'ensemble  $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\}$  avec  $n \in \mathbb{N}$ , définie, pour tout  $k \in \{0, 1, 2, \dots, (n-1), n\}$ , par :

$$\mathcal{H}_A\left(\frac{k}{n}\right) = \sup_{(x_1, \dots, x_n) \in \mathbb{R}^n: \sum_{i=1}^n \mathbb{1}_A(x_i) = k} \left\{ \min_{i \in \{1, \dots, n\}} \mu_{\mathcal{X}_i}(x_i) \right\}, \quad (\text{II.50})$$

où  $\mu_{\mathcal{X}_i}$  est la fonction d'appartenance de  $\mathcal{X}_i$  définie par (II.1).

Trutschnig a souligné dans [114] que cette construction est fortement liée à celle définie dans (II.46) et vérifie la relation suivante :

$$\forall k \in \{0, 1, 2, \dots, (n-1), n\}, \quad \mathcal{H}_A\left(\frac{k}{n}\right) = (\mathcal{H}^n(A))\left(\frac{k}{n}\right). \quad (\text{II.51})$$

### II.2.2.2 Distribution de probabilité floue

Nous présentons, dans cette partie, le concept de la distribution de probabilité floue et sa relation avec l'histogramme flou. Nous présentons deux visions pour définir une distribution de probabilité floue. La première, au sens de Trutschnig [114], voit une distribution de probabilité comme limite, quand le nombre d'observations tend vers l'infini, de l'histogramme flou définie par (II.46). La deuxième, au sens de Viertl [119, 117], considère qu'une distribution de probabilité floue est induite par une densité de probabilité floue. Cette densité de probabilité floue est définie par une intégrale floue au sens de Viertl.

#### Définition II.14 (Distribution de probabilité floue au sens de Trutschnig)

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et soit  $\mathcal{B}(\Omega)$  une tribu borélienne sur  $\Omega$ . Soit  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  une variable aléatoire floue, pour tout  $A \in \mathcal{B}(\Omega)$ , le nombre flou  $\mathcal{P}(A)$  est appelé probabilité à valeurs floues de  $A$  induite par  $\mathcal{X}$  ou, plus formellement, l'application  $\mathcal{P} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  est appelée distribution de probabilité floue induite par  $\mathcal{X}$ , si elle vérifie :

1.  $\text{supp}(\mathcal{P}(A)) \subseteq [0, 1]$
2.  $\mathcal{P}(\Omega) = \mathbb{1}_{\{1\}}$ ,  $\mathcal{P}(\emptyset) = \mathbb{1}_{\{0\}}$
3. Si  $A \subseteq B \subseteq \Omega$ , alors  $\mathcal{P}(A) \preceq \mathcal{P}(B)$
4. Si  $A \cap B = \emptyset$ , alors  $\mathcal{P}(A \cup B) \subseteq \mathcal{P}(A) \oplus \mathcal{P}(B)$
5.  $\mathcal{P}(A^c) = \mathbb{1}_{\{1\}} \ominus \mathcal{P}(A)$ .

En se basant, d'une part, sur la définition de la fréquence relative floue  $\mathcal{H}^n$  (II.46) et, d'autre part, sur la définition de la distribution de probabilité floue  $\mathcal{P}$  (Définition II.14), Trutschnig a montré dans [114] la convergence uniforme de  $\mathcal{H}^n(A)$  vers  $\mathcal{P}(A)$ ,  $\forall A \in \mathcal{B}(\Omega)$ . Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et soit  $(\mathcal{X}_n)_{n \in \mathbb{N}}$  une suite des variables aléatoires floues indépendantes et identiquement distribuées. La fréquence relative supérieure et inférieure de niveau  $\alpha$ , pour tout  $A \in \mathcal{B}(\Omega)$  et tout  $x \in \Omega$ , est définie respectivement par :

$$\underline{h}_\alpha^n(A, x) = \frac{1}{n} \left| \{i : \mathcal{X}_{i\alpha}(x) \subseteq A\} \right|, \quad (\text{II.52})$$

et

$$\overline{h}_\alpha^n(A, x) = \frac{1}{n} \left| \{i : \mathcal{X}_{i\alpha}(x) \cap A \neq \emptyset\} \right|. \quad (\text{II.53})$$

La convergence uniforme, quand le nombre d'observations floues tends vers l'infini, de  $\mathcal{H}^n$  vers  $\mathcal{P}$  est donnée par le théorème suivant [114] :

**Théorème II.15 (Convergence de l'histogramme floue)** *Soient  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $\mathcal{B}(\Omega)$  une tribu borélienne sur  $\Omega$  et soit  $A \in \mathcal{B}(\Omega)$ . Soient  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$   $n$  variables aléatoires floues indépendantes et identiquement distribuées, alors, il existe un ensemble  $\mathfrak{A} \in \mathcal{A}$ , vérifie  $P(\mathfrak{A}) = 0$ , et tel que pour tout  $x \in \mathfrak{A}^c$ , on a :*

$$\lim_{n \rightarrow \infty} d_{H,\infty}(\mathcal{H}^n(A, x), \mathcal{P}(A)) = 0, \quad (\text{II.54})$$

où  $d_{H,\infty}$  est la distance définie en (II.10).

Donc, pour une suite des variables aléatoires floues  $(\mathcal{X}_n)_{n \in \mathbb{N}}$ , la fréquence relative floue  $\mathcal{H}^n(\cdot)$  est un estimateur fortement convergent de la distribution de probabilité floue correspondante  $\mathcal{P}(\cdot)$  [114].

Autrement vu, Viertl [119, 117] définit une distribution de probabilité floue via une densité de probabilité floue. Les densités de probabilités floues sont, d'après Viertl [117], les fonctions induites par l'histogramme flou qui sont à valeurs floues.

**Définition II.16 (Intégrale floue au sens de Viertl [117])** *Soit  $\mathfrak{f} : \Omega \rightarrow \mathcal{F}_c(\Omega)$  une fonction à valeurs floues, c'est-à-dire  $\forall x \in \Omega$ ,  $\mathfrak{f}(x)$  est un nombre flou dont la coupe de niveau  $\alpha$  est notée  $\mathfrak{f}_\alpha(x) = [\underline{f}_\alpha(x), \overline{f}_\alpha(x)]$ . L'intégrale floue de  $\mathfrak{f}$  notée  $\mathfrak{I}$  :*

$$\mathfrak{I} := \int_{\Omega} \mathfrak{f}(x) dx, \quad (\text{II.55})$$

est une quantité floue définie par ses coupes de niveau, notées  $\forall \alpha \in (0, 1]$ ,  $\mathfrak{I}_\alpha = [\underline{I}_\alpha, \overline{I}_\alpha]$ , avec :

$$\underline{I}_\alpha = \int_{\Omega} \underline{f}_\alpha(x) dx \text{ et } \overline{I}_\alpha = \int_{\Omega} \overline{f}_\alpha(x) dx. \quad (\text{II.56})$$

**Définition II.17 (Densité de probabilité floue au sens de Viertl [117])** *Soit  $(\Omega, \mathcal{A})$  un espace mesurable, les densités de probabilités floues  $\mathfrak{f}(\cdot)$  sur  $(\Omega, \mathcal{A})$  sont les fonctions à valeurs floues dont l'intégrale floue est  $\mathfrak{I}$  avec  $1 \in \mathfrak{I}_1$  (c'est-à-dire avec des densités  $f$  appartenant à  $\mathfrak{I}_1$  telles que  $\int_{\Omega} f(x) dx = 1$ ).*

La distribution de probabilité floue, d'après Viertl [117], est définie par une procédure d'intégrale floue appelée  $(P)$ -intégrale notée  $(P) \int_A \mathfrak{f}(x) dx$ ,  $\forall A \subseteq \mathcal{A}$ .

**Définition II.18 (Distribution de probabilité floue au sens de Viertl [117])** Soit  $(\Omega, \mathcal{A})$  un espace mesurable et soit  $f$  une densité de probabilité floue connue (Définition II.17). Cette densité floue induit une distribution de probabilité floue, notée  $\mathcal{P}$ , est définie par ses coupes de niveau  $\alpha$ , notées  $\mathcal{P}_\alpha = [\underline{p}_\alpha, \bar{p}_\alpha]$ , avec  $\forall A \subseteq \mathcal{A}$  et  $\alpha \in (0, 1]$  :

$$\underline{p}_\alpha(A) := \inf \left\{ \int_A f(x) dx : f \in \mathcal{F}(\alpha) \right\}, \quad (\text{II.57})$$

et

$$\bar{p}_\alpha(A) := \sup \left\{ \int_A f(x) dx : f \in \mathcal{F}(\alpha) \right\}, \quad (\text{II.58})$$

où  $\mathcal{F}(\alpha)$  est l'ensemble des densités de probabilités défini par :

$$\mathcal{F}(\alpha) = \left\{ f, \int_\Omega f(x) dx = 1 : \underline{f}_\alpha(x) \leq f(x) \leq \bar{f}_\alpha(x) \right\}. \quad (\text{II.59})$$

La distribution de probabilité floue  $\mathcal{P}$  (Définition II.18) vérifie, pour tout  $A, B \subseteq \Omega$  telles que  $A \cap B = \emptyset$  et  $\alpha \in (0, 1]$ , les deux inégalités suivantes [117] :

$$\underline{p}_\alpha(A \cup B) \geq \underline{p}_\alpha(A) + \underline{p}_\alpha(B). \quad (\text{II.60})$$

$$\bar{p}_\alpha(A \cup B) \leq \bar{p}_\alpha(A) + \bar{p}_\alpha(B), \quad (\text{II.61})$$

Dans ce cas, la coupe de niveau  $\underline{p}_\alpha$  (rsp.  $\bar{p}_\alpha$ ) peut être vue comme une capacité convexe (II.13) (rsp. concave (II.14)).

### II.2.3 Estimation floue de densité

Une variable aléatoire floue induit une densité de probabilité floue. Reste maintenant le problème classique de son estimation à partir d'un ensemble fini d'observations. Dans cette section, nous nous intéressons à quelques méthodes pertinentes de la littérature permettant de donner une estimation floue de la densité  $f$  sous-jacente à un ensemble fini d'observations floues. Les méthodes que nous présentons dans cette section sont des extensions de méthodes classiques d'estimation non-paramétrique précise, présentées en section I.2, de la densité sous-jacente à un ensemble fini d'observations réelles précises. On doit ces extensions aux travaux de Arefi et al. [2] utilisant la décomposition par coupe de niveau pour généraliser, aux variables aléatoires floues, les estimateurs de densité classiques.

La première extension proposée par Arefi et al. [2] est celle s'appuyant sur l'estimation de la densité par histogramme (cf. section I.2.1.3). Rappelons que, pour un histogramme classique, l'estimation de la densité de probabilité, considérant un ensemble fini d'observation précise  $(x_1, \dots, x_n)$ , s'obtient en deux étapes. La première consiste à subdiviser l'intervalle de référence  $\Omega \subseteq \mathbb{R}$ , en une partition de  $p$  sous ensembles de pas  $\Delta = \frac{|\Omega|}{p}$  (c'est-à-dire,  $\forall k \in \{1, \dots, p\}$ ,  $\Omega = \bigcup_{k=1}^p A_k$ ,  $A_i \cap A_j = \emptyset, \forall i \neq j$  et  $|A_k| = \Delta$ .) et à associer, à chaque sous-ensemble, de la partition, un accumulateur  $\text{Acc}_k$  donné par (I.2) :

$$\text{Acc}_k = \sum_{i=1}^n \mathbb{1}_{A_k}(x_i),$$

où  $\mathbb{1}_A(\cdot)$  est la fonction indicatrice de  $A$  définie par (I.3).

La seconde étape consiste à estimer la densité en chaque point  $x \in \Omega$  (I.6) :

$$\widehat{f}_{\Delta}^n(x) = \frac{1}{n\Delta} \sum_{k=1}^p \text{Acc}_k \mathbb{1}_{A_k}(x).$$

L'extension proposée par Arefi et al. [2] de la méthode d'estimation par histogramme de la densité de probabilité est définie comme suit :

**Définition II.19 (Extension de la méthode d'histogramme)** Soit  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  un ensemble fini de nombres flous issus d'une même loi de probabilité de densité  $f$  à support bornée  $\Omega$ . On suppose  $\Omega = [e_{\min}, e_{\max}]$  et on définit une partition de  $\Omega$  telle que  $e_{\min} = w_1, w_1 + \Delta, \dots, w_1 + (p-1)\Delta = e_{\max}$ . L'extension de la méthode d'estimation par histogramme est définie par ses densités floues  $\widehat{f}_{\alpha}^n$  dont la coupe de niveau est donnée,  $\forall \alpha \in (0, 1]$ , par :

$$\begin{cases} \widehat{f}_{\alpha}^n(x) = q_{k\alpha}, & \text{pour } w_k \leq x < w_k + \Delta, \forall k \in \{1, \dots, p\}, \\ \widehat{f}_{\alpha}^n(e_{\max}) = q_{p\alpha}, \\ \widehat{f}_{\alpha}^n(x) = 0, & \text{sinon} \end{cases} \quad (\text{II.62})$$

où

$$\begin{aligned} q_{k\alpha} &= \frac{1}{n\Delta} \sum_{i=1}^n \left[ \inf \left( \mathbb{1}_{[w_k, w_k + \Delta]}(\chi); \chi \in \mathcal{X}_{i\alpha} \right), \sup \left( \mathbb{1}_{[w_k, w_k + \Delta]}(\chi); \chi \in \mathcal{X}_{i\alpha} \right) \right] \\ &= \left[ \frac{1}{n\Delta} \sum_{i=1}^n \inf \left( \mathbb{1}_{[w_k, w_k + \Delta]}(\chi); \chi \in \mathcal{X}_{i\alpha} \right), \frac{1}{n\Delta} \sum_{i=1}^n \sup \left( \mathbb{1}_{[w_k, w_k + \Delta]}(\chi); \chi \in \mathcal{X}_{i\alpha} \right) \right]. \end{aligned}$$

La deuxième extension proposée par Arefi et al. [2] est celle de l'estimateur simple de densité de probabilité (cf. section I.2.2). Rappelons que l'estimateur simple de la densité en un point  $x$  est obtenu en comptant le nombre d'observations dans un intervalle autour de  $x$ . Cet estimateur est donné (I.14) par :

$$\widehat{f}_{\Delta}^n(x) = \frac{1}{2n\Delta} \sum_{i=1}^n \mathbb{1}_{[x_i - \Delta, x_i + \Delta]}(x).$$

L'extension de l'estimateur simple de la densité, proposée par Arefi et al. [2], est définie comme suit :

**Définition II.20 (Extension de l'estimateur simple)** Soit  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  un ensemble fini de nombres flous, l'extension de l'estimateur simple est définie par des densités floues  $\widehat{f}_{\alpha}^n$  dont la coupe de niveau est donnée,  $\forall \alpha \in (0, 1]$ , par :

$$\begin{aligned} \widehat{f}_{\alpha}^n(x) &= \frac{1}{2n\Delta} \sum_{i=1}^n \left( \inf \left( \mathbb{1}_{(x-\Delta, x+\Delta]}(\chi); \chi \in \mathcal{X}_{i\alpha} \right), \sup \left( \mathbb{1}_{(x-\Delta, x+\Delta]}(\chi); \chi \in \mathcal{X}_{i\alpha} \right) \right) \\ &= \left( \frac{1}{2n\Delta} \sum_{i=1}^n \inf \left( \mathbb{1}_{(x-\Delta, x+\Delta]}(\chi); \chi \in \mathcal{X}_{i\alpha} \right), \frac{1}{2n\Delta} \sum_{i=1}^n \sup \left( \mathbb{1}_{(x-\Delta, x+\Delta]}(\chi); \chi \in \mathcal{X}_{i\alpha} \right) \right). \end{aligned}$$

Notons que la coupe de niveau  $\alpha$  de  $\widehat{f}^n$  est un intervalle donné,  $\forall x \in \Omega$ , par :  $\widehat{f}_\alpha^n(x) = [\underline{\widehat{f}}_\alpha^n(x), \overline{\widehat{f}}_\alpha^n(x)]$ . Arefi et al. ont montré dans [2] la convergence de cet estimateur vers la vraie densité  $f$ , en terme de l'espérance et de la variance :

$$\text{Si } \Delta \rightarrow 0 \text{ quand } n \rightarrow \infty, \text{ alors } \mathbb{E}(\underline{\widehat{f}}_\alpha^n) \text{ (ou } \mathbb{E}(\overline{\widehat{f}}_\alpha^n)) \rightarrow f,$$

et

$$\text{si } \Delta \rightarrow 0 \text{ quand } n\Delta \rightarrow \infty, \text{ alors } \text{var}(\underline{\widehat{f}}_\alpha^n) \text{ (ou } \text{var}(\overline{\widehat{f}}_\alpha^n)) \rightarrow 0.$$

La troisième extension proposée par Arefi et al. [2] est celle de l'estimation par noyau de Parzen-Rosenblatt (cf. section I.2.3). Rappelons que l'estimateur de densité à noyau en un point  $x \in \Omega$  consiste à construire une densité comme étant une moyenne d'évaluation ponctuelle en  $x$  de noyaux sommatifs (I.8), chacun de ces noyaux étant centré sur une des données observées  $x_i$ . Cet estimateur est donné (I.24) par :

$$\forall x \in \Omega, \widehat{f}_{\kappa_\Delta}^n(x) = \frac{1}{n\Delta} \sum_{i=1}^n \kappa\left(\frac{x - x_i}{\Delta}\right).$$

où  $\kappa$  est un noyau sommatif et  $\Delta > 0$  la largeur de bande de l'estimateur.

L'extension proposée par Arefi et al. [2] de cet estimateur est définie comme suit :

**Définition II.21 (Extension de l'estimateur à noyau)** Soit  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  un ensemble fini de  $n$  nombres flous, l'extension de l'estimateur à noyau de Parzen-Rosenblatt est définie par les densités floues  $\widehat{f}_\kappa^n$  dont la coupe de niveau est donnée,  $\forall \alpha \in (0, 1]$ , par :

$$\begin{aligned} \widehat{f}_{\kappa\alpha}^n(x) &= \frac{1}{n\Delta} \sum_{i=1}^n \kappa\left(\frac{x - \mathcal{X}_{i\alpha}}{\Delta}\right) \\ &= \frac{1}{n\Delta} \sum_{i=1}^n \left( \inf\left(\kappa\left(\frac{x - \chi}{\Delta}\right); \chi \in \mathcal{X}_{i\alpha}\right), \sup\left(\kappa\left(\frac{x - \chi}{\Delta}\right); \chi \in \mathcal{X}_{i\alpha}\right) \right) \\ &= \left( \frac{1}{n\Delta} \sum_{i=1}^n \inf\left(\kappa\left(\frac{x - \chi}{\Delta}\right); \chi \in \mathcal{X}_{i\alpha}\right), \frac{1}{n\Delta} \sum_{i=1}^n \sup\left(\kappa\left(\frac{x - \chi}{\Delta}\right); \chi \in \mathcal{X}_{i\alpha}\right) \right). \end{aligned}$$

La convergence de cet estimateur vers la vraie densité  $f$ , bornée et continue sur chaque  $x \in \Omega$ , en terme de biais et de variance, a été démontrée par Arefi et al. [2] via le théorème suivant :

**Théorème II.22 (Convergence de l'extension de l'estimateur à noyau)** Soit  $\mathcal{X}$  une variable aléatoire floue (Définition II.7) et soit  $\widehat{f}_\kappa^n$  l'extension de l'estimateur à noyau de Parzen-Rosenblatt (Définition II.21). On a :

1. Si  $\Delta \rightarrow 0$  quand  $n \rightarrow \infty$ , alors biais  $(\underline{\widehat{f}}_{\kappa\alpha}^n)$  (ou Viais  $(\overline{\widehat{f}}_{\kappa\alpha}^n)$ )  $\rightarrow 0$ .
2. Si  $\int_{-\infty}^{+\infty} \left( \inf(\kappa(u) : (x - \Delta u) \in \mathcal{X}_{1\alpha}) \right)^2 du < \infty$ ,  $\Delta \rightarrow 0$  et  $n\Delta \rightarrow \infty$  quand  $n \rightarrow \infty$ , alors

$$\text{var}(\underline{\widehat{f}}_\alpha^n) \text{ (ou } \text{var}(\overline{\widehat{f}}_\alpha^n)) \rightarrow 0.$$

## II.3 Mesure de probabilité imprécise et données précise

Il peut exister des cas où la connaissance que l'on a de la probabilité sous-jacente à un ensemble fini d'observations est imprécise alors que les observations sont précises. En effet, en estimation non-paramétrique, l'estimation de la densité en un point revient à estimer cette densité dans le voisinage de ce point. Par exemple, avec la méthode d'estimation de Parzen-Rosenblatt, présentée en section I.2.3, le noyau sommatif  $\kappa$  (I.8) détermine la forme du voisinage autour de point  $x \in \Omega$  et sa largeur de bande  $\Delta$  contrôle la taille de ce voisinage, c'est-à-dire le nombre d'observations prises pour effectuer la moyenne locale.

Dans la plupart du temps, le voisinage nécessaire à l'estimation est mal défini ou mal connu. Par exemple, avec la méthode d'estimation par noyau, le noyau d'Epanechnikov (I.32) est le plus utilisé parce que ce noyau minimise le critère AMISE (la distance entre la densité estimée et la vraie densité). Ce noyau peut être changé, en utilisant un autre critère de minimisation comme l'ont proposé Devroy et Lugosi [27, 28].

Généralement, dans beaucoup de domaines d'applications, les noyaux sommatifs et leur largeur de bande sont choisis de façon très empirique. Le caractère empirique du choix du noyau pouvait être pris en compte, d'après [72], en remplaçant un noyau par une famille de noyaux. Cette famille de noyaux peut être représentée par un noyau non-sommatif (ou noyau flou) appelé *noyau maxitif*. Un noyau maxitif représente l'ensemble des noyaux sommatifs qu'il domine [72].

Dans beaucoup de domaines d'applications, généralement, les noyaux sommatifs et leur largeur de bande sont choisis de façon très empirique. Comme le proposent Loquin et Strauss [72] le caractère empirique du choix du noyau pouvait être pris en compte en remplaçant un noyau par une famille de noyaux. Cette famille de noyaux peut être représentée par un noyau non-sommatif (ou noyau flou) appelé *noyau maxitif*. Un noyau maxitif représente l'ensemble des noyaux sommatifs qu'il domine [72]. En s'appuyant sur ces types de noyaux non-sommatifs, Loquin et Strauss proposent dans [72] d'étendre l'opérateur espérance sommatif (I.9), en parlant de l'*espérance maxitive*. Cette espérance forme la base de l'*estimation maxitive de la fonction de répartition* proposée par Loquin et Strauss [71].

Nous présentons, dans cette section, la définition des noyaux maxitifs ainsi que leurs propriétés. Nous introduisons ensuite l'opérateur d'espérance maxitive basé sur un noyau maxitif. Nous nous concentrons sur le principe et la définition de l'estimation imprécise de la fonction de répartition ainsi que sur ses propriétés. Cette estimation représente une étape majeure dans la construction de l'extension maxitive de la densité de probabilité que nous proposons. Nous abordons, à la fin de cette section, la question de l'existence d'une extension maxitive de la densité sous-jacente à un ensemble fini d'observations.

### II.3.1 Noyaux maxitifs

**Définition II.23 (Noyau maxitif [72])** *Un noyau maxitif est une fonction  $\pi : \Omega \rightarrow [0, 1]$ , vérifiant la propriété de maxitivité :*

$$\sup_{u \in \Omega} \pi(u) = 1. \quad (\text{II.63})$$

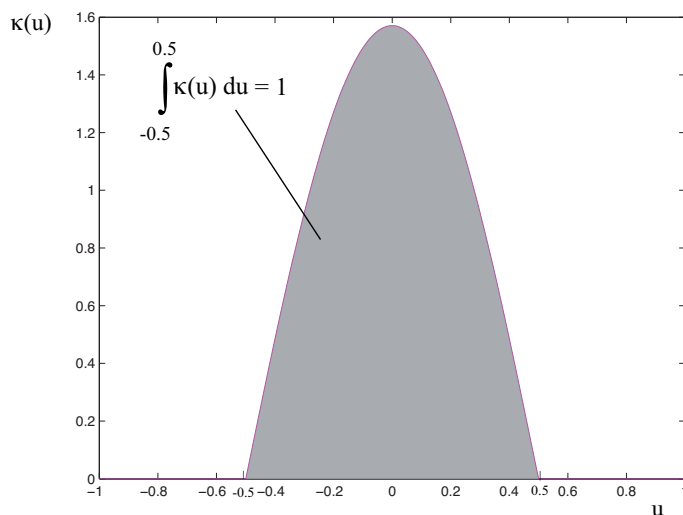
On peut définir, à partir d'un noyau maxitif  $\pi$ , un noyau maxitif  $\pi_\Delta^x$  translaté en  $x \in \Omega$

et dilaté avec une largeur de bande  $\Delta > 0$ , par :

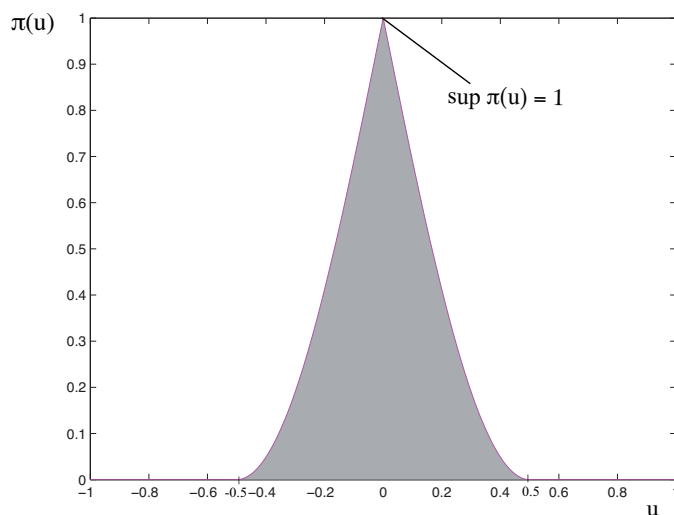
$$\forall u \in \Omega, \pi_{\Delta}^x(u) = \pi\left(\frac{u-x}{\Delta}\right). \quad (\text{II.64})$$

Par convention,  $\pi(u) = \pi_1^0(u)$ .

La Figure II.1 illustre la différence entre un noyau sommatif  $\kappa$  (Figure II.1 (a)) et un noyau maxitif  $\pi$  (Figure II.1 (b)).



(a) Noyau sommatif.



(b) Noyau maxitif.

**Figure II.1 :** Un noyau sommatif  $\kappa$  vérifie la propriété de sommativité (a) et un noyau maxitif  $\pi$  vérifie la propriété de maxitivité (b).

Un noyau maxitif peut être vu [72] comme une distribution de possibilité induisant une mesure de possibilité [35, 19] notée  $\Pi_{\pi}$  et définie, d'après l'expression (II.36), par :

$$\forall A \subseteq \Omega, \Pi_{\pi}(A) = \sup_{u \in A} \pi(u). \quad (\text{II.65})$$

De façon dual, le noyau maxitif  $\pi$  induit [72] une mesure de nécessité, notée  $N_\pi$  et définie, d'après l'expression (II.38), par :

$$\forall A \subseteq \Omega, N_\pi(A) = 1 - \Pi_\pi(A^c), \quad (\text{II.66})$$

où  $A^c$  est le complémentaire de  $A$  dans  $\Omega$ .

On dit qu'un noyau maxitif  $\pi$  domine [72] un noyau sommatif  $\kappa$ , si la mesure de possibilité  $\Pi_\pi$  domine la mesure de probabilité  $P_\kappa$ , c'est-à-dire :

$$\forall A \subseteq \Omega, P_\kappa(A) \leq \Pi_\pi(A). \quad (\text{II.67})$$

En ce sens, un noyau maxitif définit l'ensemble des noyaux sommatifs qu'il domine [72]. Cet ensemble est noté  $\mathcal{M}(\pi)$ , défini par :

$$\mathcal{M}(\pi) = \left\{ \kappa \in \mathcal{K}(\Omega) : \forall A \subseteq \Omega, N_\pi(A) \leq P_\kappa(A) \leq \Pi_\pi(A) \right\}. \quad (\text{II.68})$$

Loquin et Strauss ont montré dans [72] qu'il existe une relation entre noyaux sommatifs et noyaux maxitifs. Cette relation s'appuie sur la transformation probabilité-possibilité dite *objective* de Dubois et al. [37, 38]. Cette transformation permet de définir le noyau maxitif le plus spécifique qui domine le noyau sommatif  $\kappa$  et vérifie les conditions suivantes :

1. Le respect de l'ordre, c'est-à-dire  $\forall u, v \in \Omega, \kappa(u) \geq \kappa(v) \Rightarrow \pi(u) \geq \pi(v)$ .
2. La domination, c'est-à-dire  $\kappa \in \mathcal{M}(\pi)$ .

**Définition II.24 (Noyau maxitif le plus spécifique [72])** Soit  $\kappa$  un noyau sommatif, le noyau maxitif le plus spécifique qui domine  $\kappa$ , noté  $\pi_{[\kappa]}$ , obtenu par une transformation probabilité-possibilité objective de  $\kappa$ , est défini par :

$$\forall u \in \Omega, \pi_{[\kappa]}(u) = 1 - P_\kappa(I_u), \quad (\text{II.69})$$

où  $I_u$  est l'intervalle de confiance de niveau  $P_\kappa(I_u)$  défini par :

$$I_u = \{x \in \Omega, \kappa(x) \geq \kappa(u)\}. \quad (\text{II.70})$$

Dans ce contexte, Dubois et al. ont montré dans [37] que le noyau triangulaire  $T_\Delta^x$  de support  $[x - \Delta, x + \Delta]$  est le noyau maxitif le plus spécifique qui domine tous les noyaux sommatifs monomodaux et symétriques de support  $[x - \Delta, x + \Delta]$  [69].

Dans le même contexte, Dubois et al. ont défini dans [39] une autre transformation probabilité-possibilité dite *subjective*. Cette transformation permet de définir un noyau maxitif moins spécifique qui domine  $\kappa$  et vérifie les mêmes conditions, c'est-à-dire le respect de l'ordre et de la domination [69].

**Définition II.25 (Noyau maxitif moins spécifique [72])** Soit  $\kappa$  un noyau sommatif, le noyau maxitif moins spécifique qui domine  $\kappa$ , noté  $\pi_{(\kappa)}$ , obtenu par la transformation probabilité-possibilité subjective de  $\kappa$  est défini par :

$$\forall u \in \Omega, \pi_{(\kappa)}(u) = \int_{\Omega} \min(\kappa(x), \kappa(u)) dx. \quad (\text{II.71})$$



Un résultat intéressant a été montré dans ([72], théorème 7) concernant la spécificité du noyau maxitif. En effet, pour un noyau sommatif  $\kappa$ , le noyau maxitif  $\pi_{[\kappa]}$  (II.69) est plus spécifique que le noyau maxitif  $\pi_{(\kappa)}$  (II.25) et on a :

$$\forall u \in \Omega, \pi_{(\kappa)}(u) = \pi_{[\kappa]}(u) + \lambda(I_u)\kappa(u), \quad (\text{II.72})$$

où  $\lambda$  est la mesure de Lebesgue.

La spécificité d'un noyau maxitif est définie, au sens de Loquin-Strauss [72], comme la capacité du noyau maxitif à être concentré sur un ensemble de longueur minimale. La longueur est prise au sens de la mesure de Lebesgue ([69], pages 69-70). La précision d'un noyau maxitif est proportionnelle à sa spécificité c'est-à-dire plus il est spécifique, plus il est précis ([69], page 70). Le degré de l'imprécision d'un noyau maxitif est mesuré par la *granularité* de ce noyau ([69], page 81). La granularité d'un noyau maxitif  $\pi$ , notée  $\rho(\pi)$ , est définie au sens de Loquin-Strauss [72] par :

$$\rho(\pi) = \int_{\Omega} \pi(u) du. \quad (\text{II.73})$$

Il est important de noter que la granularité d'un noyau maxitif est équivalente à la cardinalité d'un sous-ensemble flou [35, 125] ([69], page 81). La granularité, définie par (II.73), mesure à quel point le noyau maxitif est peu concentré, peu spécifique. Cette grandeur est un indice de non-spécificité du noyau maxitif ([69], page 81). En ce sens, on dit qu'un noyau maxitif  $\pi$  est *plus spécifique* qu'un noyau maxitif  $\pi$ , si  $\rho(\pi) \leq \rho(\pi)$ .

**Lemme II.26 (Granularité du noyau maxitif dilaté)** Soit  $\pi_{\Delta}$  un noyau maxitif dilaté avec une largeur de bande  $\Delta > 0$ . On a :

$$\rho(\pi_{\Delta}) = \Delta \rho(\pi), \quad (\text{II.74})$$

où  $\rho(\pi)$  est la granularité du noyau maxitif  $\pi$  définie par (II.73).

**Preuve :** On a

$$\begin{aligned} \rho(\pi_{\Delta}) &= \int_{\Omega} \pi_{\Delta}(u) du, \text{ par l'expression (II.64)} \\ &= \int_{\Omega} \pi\left(\frac{u}{\Delta}\right) du \\ &= \Delta \int_{\Omega} \pi(u) du \\ &= \Delta \rho(\pi). \end{aligned}$$

□

D'après le Lemme II.26, on a la propriété suivante :

**Propriété II.27.** Soient  $\pi$  un noyau maxitif et  $\pi^x$  un noyau maxitif translaté en  $x \in \Omega$  et soit  $\rho(\pi)$  (resp.  $\rho(\pi^x)$ ) la granularité de  $\pi$  (resp.  $\pi^x$ ). La granularité d'un noyau maxitif translaté est égale à la granularité du noyau maxitif non-translaté.

### II.3.2 Espérance mathématique maxitive

Une propriété fondamentale des noyaux maxitifs, découlant des travaux de Schmeidler [96] et Denneberg [22] d'une part et Rico et Strauss [92] d'autre part, permet de répercuter la propriété de domination des noyaux sur les estimations. En effet, si  $s$  est une fonction bornée de  $L(\Omega)$ ,  $\pi$  un noyau maxitif (Définition II.23) et  $\kappa$  un noyau sommatif (Définition I.8) dominé par  $\pi$ , on a la relation suivante [71] :

$$\mathbb{C}_{N_\pi}(s) \leq \mathbb{E}_\kappa(s) \leq \mathbb{C}_{\Pi_\pi}(s), \quad (\text{II.75})$$

où  $\mathbb{C}_{N_\pi}(s)$  (resp.  $\mathbb{C}_{\Pi_\pi}(s)$ ) est l'intégrale de Choquet de la fonction  $s$  par rapport à la mesure de nécessité  $N_\pi$  (resp. possibilité  $\Pi_\pi$ ) et  $\mathbb{E}_\kappa(s)$  est l'espérance sommative de  $s$  basée sur  $\kappa$  (Définition I.6).

Les deux intégrales de Choquet données dans l'expression (II.75) ont permis à Loquin et Strauss [71] de définir une extension de l'opérateur espérance sommative (précise) dont le voisinage est défini par le noyau maxitif  $\pi$ .

**Définition II.28 (Espérance maxitive au sens de Loquin-Strauss)** *Soit  $\pi$  un noyau maxitif et soit  $s$  une fonction bornée de  $L(\Omega)$ . Soit  $\mathcal{M}(\pi)$  l'ensemble des noyaux sommatifs définis par (II.68), l'espérance maxitive de  $s$  basée sur  $\pi$  est définie par :*

$$\overline{\mathbb{E}}_\pi(s) = [\underline{\mathbb{E}}_\pi(s), \overline{\mathbb{E}}_\pi(s)] = \{\mathbb{E}_\kappa(s) : \kappa \in \mathcal{M}(\pi)\}, \quad (\text{II.76})$$

où  $\underline{\mathbb{E}}_\pi(s) = \mathbb{C}_{N_\pi}(s)$  (resp.  $\overline{\mathbb{E}}_\pi(s) = \mathbb{C}_{\Pi_\pi}(s)$ ) est la borne inférieure (resp. supérieure) de  $\overline{\mathbb{E}}_\pi(s)$  et  $\mathbb{E}_\kappa(\cdot)$  est l'espérance sommative basée sur le noyau sommatif  $\kappa$ .

L'estimateur intervalliste  $\overline{\mathbb{E}}_\pi(s)$  représente l'ensemble des valeurs  $\mathbb{E}_\kappa(s)$  obtenues pour tous les noyaux sommatifs  $\kappa \in \mathcal{M}(\pi)$ . Loquin et Strauss ont montré dans [71] que cet opérateur vérifie la propriété de domination suivante :

**Propriété II.29 (Propriété de domination)** *Soit  $\pi$  un noyau maxitif et soit  $\mathcal{M}(\pi)$  l'ensemble des noyaux sommatifs dominés par  $\pi$  défini par (II.68). Soit  $s$  une fonction bornée de  $L(\Omega)$ , on a :*

$$\forall y \in \overline{\mathbb{E}}_\pi(s), \exists \kappa \in \mathcal{M}(\pi) : \mathbb{E}_\kappa(s) = y, \quad (\text{II.77})$$

et

$$\forall \kappa \in \mathcal{M}(\pi), \mathbb{E}_\kappa(s) \in \overline{\mathbb{E}}_\pi(s), \quad (\text{II.78})$$

où  $\mathbb{E}_\kappa(s)$  (resp.  $\overline{\mathbb{E}}_\pi(s)$ ) est l'espérance sommative (resp. maxitive) de  $s$  basée sur  $\kappa$  (resp.  $\pi$ ).

### II.3.3 Estimation imprécise de la fonction de répartition

Un noyau maxitif représente l'ensemble des noyaux sommatifs qu'il domine [72]. Cette idée a été utilisée par Loquin et Strauss [71] pour construire un *intervalle d'estimation de la fonction de répartition* en considérant un ensemble fini d'observations précises. C'est sur cette idée que nous nous appuyons pour construire un *intervalle d'estimation de la densité de probabilité* sous-jacente à un ensemble fini d'observations précises, comme nous allons montrer dans le chapitre suivant. Comme spécifie dans [71], cet intervalle contient l'ensemble des estimateurs de Parzen-Rosenblatt de la fonction de répartition estimés avec des noyaux appartenant à l'ensemble des noyaux sommatifs dominés par le noyau maxitif choisi. C'est d'ailleurs le plus spécifique ayant cette propriété.

Cette section présente le principe de l'estimation maxitive de la fonction de répartition ainsi ses propriétés.

### II.3.3.1 Principe et définition

Soit un ensemble fini d'observations  $(x_1, \dots, x_n)$  de  $n$  variables aléatoires  $(X_1, \dots, X_n)$  i.i.d de même densité de probabilité. Rappelons que l'estimateur à noyau  $\kappa$  de Parzen-Rosenblatt de la fonction de répartition  $\widehat{F}_\kappa^n$  en chaque point  $x \in \Omega$  est donné (I.48) par :

$$\widehat{F}_\kappa^n(x) = \frac{1}{n} \sum_{i=1}^n \kappa(x - x_i),$$

où  $\kappa(x) = \int_{-\infty}^x \kappa(u) du$ .

Loquin et Strauss ont montré dans [71] que cet estimateur peut s'écrire, en terme de l'espérance sommative (Définition I.9), comme suit :

$$\widehat{F}_\kappa^n(x) = \mathbb{E}_{\kappa^x}(E_n), \quad (\text{II.79})$$

où  $\mathbb{E}_\kappa(\cdot)$  est l'espérance sommative basée sur le noyau sommatif  $\kappa$  définie par (I.6) et  $E_n$  est la fonction de répartition empirique donnée (I.39) par :

$$E_n(x) = \frac{1}{n} \sum_{i=1}^n H(x - x_i),$$

où  $H$  est la fonction de Heaviside donnée (I.40) par :

$$H(x) = \begin{cases} 1 & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Loquin et Strauss ont proposé dans [71] une extension de la méthode d'estimation de la fonction de répartition de Parzen-Rosenblatt (II.79). Cette estimation imprécise (maxitive) est basée sur le noyau maxitif  $\pi$  et définie par :

$$\forall x \in \Omega, \overline{F}_\pi^n(x) = [F_\pi^n(x), \overline{F}_\pi^n(x)] = \overline{\mathbb{E}}_{\pi^x}(E_n) = [\underline{\mathbb{E}}_{\pi^x}(E_n), \overline{\mathbb{E}}_{\pi^x}(E_n)]. \quad (\text{II.80})$$

Le calcul de  $\overline{F}_\pi^n$  fait intervenir deux intégrales de Choquet. Ce calcul est donné, pour tout  $x \in \Omega$ , par :

$$\underline{\mathbb{E}}_{\pi^x}(E_n) = \mathbb{C}_{N_{\pi^x}}(E_n) = \frac{1}{n} \sum_{i=1}^n \left( (1 - \pi^x(x_i)) H(x - x_i) \right), \quad (\text{II.81})$$

$$\overline{\mathbb{E}}_{\pi^x}(E_n) = \mathbb{C}_{\Pi_{\pi^x}}(E_n) = \frac{1}{n} \sum_{i=1}^n \left( \pi^x(x_i) H(x_i - x) + H(x - x_i) \right). \quad (\text{II.82})$$

On peut trouver ce calcul en détail dans ([69], pages 110-111).

Lorsqu'il s'agit de représenter graphiquement la fonction de répartition, ou encore en vue d'une utilisation algorithmique, il est souvent nécessaire d'estimer la fonction de répartition sur un ensemble (régulièrement espacés ou non) de points de  $\Omega$ . Soient  $\{y_j\}_{j \in \{1, \dots, p\}}$  ces  $p$  points. L'algorithme de calcul de  $\{\overline{F}_\pi^n(y_j)\}_{j \in \{1, \dots, p\}}$ , en chaque point  $\{y_j\}_{j \in \{1, \dots, p\}}$  de  $\Omega$  est donné par :

**Data:** les points  $\{y_j\}_{j \in \{1, \dots, p\}}$ , les observations  $\{x_i\}_{i \in \{1, \dots, n\}}$  et le noyau maxitif  $\pi$ .

**Result:**  $\left\{ \left[ \underline{F}_\pi^n(y_j), \overline{F}_\pi^n(y_j) \right] \right\}_{j \in \{1, \dots, p\}}$ .

```

begin
  for  $j = 1$  to  $p$  do
     $\overline{F}_\pi^n(y_j) \leftarrow 0$ ,  $\underline{F}_\pi^n(y_j) \leftarrow 0$ 
    for  $i = 1$  to  $n$  do
      if  $y_j \leq x_i$  then
        |  $\overline{F}_\pi^n(y_j) = \overline{F}_\pi^n(y_j) + 1$ 
      else
        |  $\pi_i = \pi^{y_j}(x_i)$ 
        |  $\overline{F}_\pi^n(y_j) = \overline{F}_\pi^n(y_j) + \pi_i$ 
      end
      if  $y_j \geq x_i$  then
        |  $\underline{F}_\pi^n(y_j) = \underline{F}_\pi^n(y_j) + 1$ 
      else
        |  $\pi_i = \pi^{y_j}(x_i)$ 
        |  $\underline{F}_\pi^n(y_j) = \underline{F}_\pi^n(y_j) + \pi_i$ 
      end
    end
     $\underline{F}_\pi^n(y_j) = n - \overline{F}_\pi^n(y_j)$ 
  end
end

```

**Algorithm 1:** Calcul de  $\left\{ \left[ \underline{F}_\pi^n(y_j), \overline{F}_\pi^n(y_j) \right] \right\}_{j \in \{1, \dots, p\}}$ .

### II.3.3.2 Propriétés

L'intervalle d'estimation de la fonction de répartition  $\overline{F}_\pi^n$  (II.80), basé sur un noyau maxitif  $\pi$ , est le plus spécifique contenant l'ensemble d'estimation ponctuelle  $\widehat{F}_\kappa^n$ , basé sur un noyau sommatif  $\kappa$ , obtenu par la méthode de Parzen-Rosenblatt pour tout noyau sommatif  $\kappa$  dominé par le noyau maxitif  $\pi$ . D'après la propriété de domination des noyaux maxitifs (Propriété II.29), on a :

$$\forall x \in \Omega, \forall \kappa \in \mathcal{M}(\pi), \widehat{F}_\kappa^n(x) \in \overline{F}_\pi^n(x), \quad (\text{II.83})$$

et

$$\forall y \in \overline{F}_\pi^n(x), \exists \kappa \in \mathcal{M}(\pi), \widehat{F}_\kappa^n(x) = y, \quad (\text{II.84})$$

où  $\mathcal{M}(\pi)$  est l'ensemble de noyaux sommatifs dominés par  $\pi$  défini par (II.68).

Un théorème intéressant a été montré dans ([69] page 112) concernant l'imprécision de l'intervalle d'estimation de la fonction de répartition (II.80). L'intégrale de l'imprécision de  $\overline{F}_\pi^n$  est égale à la granularité du noyau maxitif  $\pi$ . En effet, soit  $\Psi_\pi^n$  l'imprécision de  $\overline{F}_\pi^n$ , définie par :

$$\forall x \in \Omega, \Psi_\pi^n(x) = \overline{F}_\pi^n(x) - \underline{F}_\pi^n(x), \quad (\text{II.85})$$

où  $\underline{F}_\pi^n(\cdot)$  (rsp.  $\overline{F}_\pi^n(\cdot)$ ) est la borne inférieure (rsp. supérieure) de  $\overline{F}_\pi^n(\cdot)$  définie par (II.81) (rsp. (II.82)).

L'intégrale de  $\Psi_\pi^n$  est donnée par :

$$\int_{\Omega} \Psi_\pi^n(x) dx = \rho(\pi), \quad (\text{II.86})$$

où  $\rho(\pi)$  est la granularité de noyau maxitif  $\pi$  définie par (II.73).

**Corollaire II.30.** *Soit  $\pi_\Delta$  un noyau maxitif dilaté avec une largeur de bande  $\Delta > 0$ . Soit  $\Psi_\pi^n$  l'imprécision de  $\underline{F}_\pi^n$  définie par (II.86) et soit  $\rho(\pi)$  la granularité du noyau maxitif  $\pi$  définie par (II.73), on a :*

$$\int_{\Omega} \Psi_{\pi_\Delta}^n(x) dx = \Delta \rho(\pi). \quad (\text{II.87})$$

**Preuve :** D'après le lemme II.26, pour tout  $\Delta > 0$ , on a :

$$\int_{\Omega} \Psi_{\pi_\Delta}^n(x) dx = \rho(\pi_\Delta) = \Delta \rho(\pi).$$

□

### II.3.4 Peut-on estimer la densité de probabilité par noyau maxitif ?

Dans la section précédente, nous avons vu que l'estimateur de Parzen-Rosenblatt pour les fonctions de répartition pouvait être étendu en utilisant un noyau maxitif. Rappelons que cette estimation imprécise est donnée (II.80) par :

$$\forall x \in \Omega, \underline{F}_\pi^n(x) = [\underline{F}_\pi^n(x), \overline{F}_\pi^n(x)] = \underline{\mathbb{E}}_{\pi^x}(E_n) = [\underline{\mathbb{E}}_{\pi^x}(E_n), \overline{\mathbb{E}}_{\pi^x}(E_n)].$$

Le fait de remplacer une estimation précise de la fonction de répartition par une estimation imprécise rapporte deux avantages principaux. Premier avantage : la méthode d'estimation est plus robuste à des mauvaises connaissances du noyau sommatif à utiliser [69]. Deuxième avantage : l'intégrale de l'imprécision de l'intervalle d'estimation dépend de la granularité du noyau maxitif utilisé et de la largeur de bande de ce noyau.

L'expression (II.80) est particulièrement simple. La question venant naturellement, suite à ces travaux, est la suivante : serait-il possible d'utiliser une extension analogue pour réaliser une estimation imprécise  $\underline{f}_\pi^n$  de  $f$  sous la forme :

$$\forall x \in \Omega, \underline{f}_\pi^n(x) = [\underline{f}_\pi^n(x), \overline{f}_\pi^n(x)] = \underline{\mathbb{E}}_{\pi^x}(e_n) = [\underline{\mathbb{E}}_{\pi^x}(e_n), \overline{\mathbb{E}}_{\pi^x}(e_n)]. \quad (\text{II.88})$$

Cependant, l'estimation imprécise de la fonction de répartition (II.80) fait intervenir deux intégrales de Choquet de la fonction  $E_n$  (voir les expressions (II.81) et (II.82)) nécessitant le fait que la fonction à intégrer soit bornée [13]. Dans le même ordre, l'intervalle d'estimation  $\underline{f}_\pi^n$  fait intervenir deux intégrales de Choquet de la mesure empirique  $e_n$ . Or  $e_n = \sum_{i=1}^n \delta^{x_i}$  est une combinaison linéaire d'impulsions de Dirac, qui est une fonction au sens de Schwartz [98]. C'est une fonction non bornée et donc l'expression (II.88) n'a pas de sens.

Dans la suite de ce manuscrit, nous montrons comment, en utilisant la *théorie de distribution* de Schwartz [98], il nous est possible de construire une estimation maxitive de la densité de probabilité sous-jacente à un ensemble fini d'observations sur un mode analogue à celle de la fonction de répartition.

# Chapitre III

## Estimation maxitive de densité de probabilité

### Introduction

Nous avons vu, au chapitre I, que l'estimateur à noyau est une des méthodes les plus populaires en estimation non-paramétrique de la densité sous-jacente à un ensemble fini d'observations. Nous avons aussi remarqué qu'une variation sur la forme de noyau conduit à un ensemble d'estimations quand le nombre d'observations est insuffisant. Dans le chapitre II, nous avons souligné qu'un noyau maxitif peut être utilisé pour représenter l'imprécision de connaissance du voisinage le plus approprié à une estimation. Cette propriété a été utilisée pour réaliser une estimation imprécise de la fonction de répartition.

Nous proposons, dans ce troisième chapitre, une extension de la méthode d'estimation par noyau de la densité sous le nom d'**estimation maxitive de la densité de probabilité**. Cette méthode d'estimation imprécise de la densité est fortement liée à celle de la fonction de répartition. En section III.1, nous présentons une reformulation pratique de l'estimateur de densité de Parzen-Rosenblatt. Cette reformulation est basée sur la *théorie de distribution de Schwartz* via la décomposition de la dérivée première du noyau utilisé. Nous présentons, en section III.2, la méthode d'estimation maxitive de densité que nous proposons. Nous présentons également un algorithme de faible complexité permettant de réaliser cette estimation. Nous étudions certaines propriétés de cet estimateur, parmi lesquelles sa convergence et son imprécision. En section III.3, nous présentons deux méthodes d'estimations plus spécifiques de la densité. Cette spécificité est due à des restrictions de la forme des noyaux utilisés. Nous donnons des algorithmes de calcul pour chacune de ces méthodes. Nous montrons les liens entre ces estimations imprécises et l'estimateur de densité de Parzen-Rosenblatt.

## III.1 Reformulation de l'estimateur de densité de Parzen-Rosenblatt

### III.1.1 Mesure empirique et estimateur de densité de Parzen-Rosenblatt

Nous avons vu en section II.3.3, que l'estimateur  $\widehat{F}_\eta^n$  de Parzen-Rosenblatt, en chaque point  $x \in \Omega$ , de la fonction de répartition peut s'écrire comme l'espérance mathématique de la fonction de répartition empirique  $E_n$  dans un voisinage probabiliste défini par le noyau sommatif  $\eta$  translaté en  $x$  :

$$\widehat{F}_\eta^n(x) = \mathbb{E}_{\eta^x}(E_n).$$

Nous proposons de montrer, de façon similaire à cette écriture, que l'estimateur de densité de Parzen-Rosenblatt  $\widehat{f}_\kappa^n$ , en chaque point  $x \in \Omega$ , peut s'écrire comme l'espérance mathématique de la mesure empirique  $e_n$  dans un voisinage probabiliste défini par le noyau sommatif  $\kappa$  translaté en  $x$  :

$$\widehat{f}_\kappa^n(x) = \mathbb{E}_{\kappa^x}(e_n).$$

Cette démonstration utilise le formalisme du produit de convolution, dont nous rappelons la définition et certaines propriétés.

**Définition III.1 (Produit de convolution ([9] page 76))** Soient  $f$  et  $g$  deux fonctions de  $L^1(\Omega)$ , leur produit de convolution noté  $(f \star g)$  est défini par :

$$\forall x \in \Omega, (f \star g)(x) = \int_{\Omega} f(u)g(x-u)du. \quad (\text{III.1})$$

Soit  $g$  une fonction de  $L^1(\Omega)$ , on appelle translaté de  $g$  en  $x \in \Omega$  la fonction  $g^x$  définie par :

$$\forall u \in \Omega, g^x(u) = g(u-x). \quad (\text{III.2})$$

Si la fonction  $g$  translatée en  $x$  est symétrique, c'est-à-dire  $\forall u \in \Omega, g(u-x) = g(x-u)$ , alors le produit de convolution peut être interprété comme le produit scalaire  $\langle \cdot, \cdot \rangle$ , défini sur  $L^1(\Omega)$ , de la fonction  $f$  avec la fonction  $g$  translatée en  $x$  et noté  $\langle f, g^x \rangle$  :

$$\forall x \in \Omega, (f \star g)(x) = \int_{\Omega} f(u)g(x-u)du = \int_{\Omega} f(u)g^x(u)du = \langle f, g^x \rangle. \quad (\text{III.3})$$

Soit  $s$  une fonction de  $L^1(\Omega)$ . Soit  $\kappa$  un noyau sommatif de  $\mathcal{K}(\Omega)$ . Le produit de convolution de  $s$  par  $\kappa$  peut être interprété comme l'espérance mathématique de la fonction  $s$  dans un voisinage probabiliste défini par le noyau sommatif  $\kappa$  translaté en  $x$  :

$$\forall x \in \Omega, (s \star \kappa)(x) = \langle s, \kappa^x \rangle = \mathbb{E}_{\kappa^x}(s). \quad (\text{III.4})$$

**Preuve :**  $\forall x \in \Omega$ , on a :

$$\begin{aligned} (s \star \kappa)(x) &= \int_{\Omega} s(u)\kappa(u-x)du \\ &= \int_{\Omega} s(u)\kappa^x(u)du \\ &= \langle s, \kappa^x \rangle \\ &= \mathbb{E}_{\kappa^x}(s). \end{aligned}$$

□

**Propriété III.2 (Mesure empirique et noyau sommatif)** *L'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_\kappa^n$ , en chaque point  $x \in \Omega$ , peut s'écrire sous la forme d'un produit scalaire d'un noyau sommatif  $\kappa$  translaté en  $x$  avec la mesure empirique  $e_n$  :*

$$\hat{f}_\kappa^n(x) = \langle e_n, \kappa^x \rangle. \quad (\text{III.5})$$

**Preuve :**  $\forall x \in \Omega$ , on a :

$$\begin{aligned} \langle e_n, \kappa^x \rangle &= \int_{\Omega} \frac{1}{n} \sum_{i=1}^n \delta^{x_i}(u) \kappa^x(u) du \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\Omega} \delta^{x_i}(u) \kappa^x(u) du \\ &= \frac{1}{n} \sum_{i=1}^n \kappa^x(x_i) \\ &= \hat{f}_\kappa^n(x). \end{aligned}$$

□

L'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_\kappa^n$ , en chaque point  $x \in \Omega$ , peut être interprété comme l'espérance mathématique de la mesure empirique  $e_n$  dans un voisinage probabiliste défini en  $x$  par le noyau sommatif  $\kappa$  :

$$\hat{f}_\kappa^n(x) = \mathbb{E}_{\kappa^x}(e_n). \quad (\text{III.6})$$

**Preuve :**  $\forall x \in \Omega$ , on a :

$$\begin{aligned} \hat{f}_\kappa^n(x) &= \langle e_n, \kappa^x \rangle \\ &= \int_{\Omega} e_n(u) \kappa^x(u) du \\ &= \mathbb{E}_{\kappa^x}(e_n). \end{aligned}$$

□

### III.1.2 Dérivation des distributions et estimateur de densité de Parzen-Rosenblatt

Notre approche est basée sur une reformulation de l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_\kappa^n$ . Cette reformulation passe par l'utilisation du lien existant entre la mesure empirique  $e_n$  et la fonction de répartition empirique  $E_n$ . Nous rappelons brièvement quelques notions de la théorie des distributions ainsi que la propriété de la dérivée au sens des distributions [98].

On doit à Schwartz [98] l'introduction de la notion de distribution. Pour définir une distribution, certaines définitions des espaces et certaines propriétés des applications doivent être définies, telles que l'espace des fonctions test, la continuité et la linéarité d'une application.



**Définition III.3 (Fonctionnelle)** Soit  $\mathcal{F}$  un espace fonctionnel (c'est-à-dire un ensemble de fonctions ayant une structure d'espace vectoriel). On définit une fonctionnelle  $T$  sur un espace fonctionnel  $\mathcal{F}$  l'application linéaire de  $\mathcal{F}$  à valeurs dans  $\mathbb{R}$  formalisée par :

$$T : \begin{cases} \mathcal{F} \rightarrow \mathbb{R} \\ s \mapsto \langle T, s \rangle. \end{cases} \quad (\text{III.7})$$

**Définition III.4 (Continuité, linéarité, dual d'un espace fonctionnel topologique)**

Soit  $T$  une fonctionnelle sur un espace fonctionnel topologique  $\mathcal{F}$ . On dit que  $T$  est continue si pour toute suite de fonctions  $(\varphi_n)_n$  de  $\mathcal{F}$  convergeant vers  $\varphi \in \mathcal{F}$ , la suite numérique  $(\langle T, \varphi_n \rangle)_n$  converge vers  $\langle T, \varphi \rangle$  ([20] page 33).

On dit que  $T$  est linéaire si pour tout réelles  $\ell_1$  et  $\ell_2$  et toutes fonctions  $\varphi_1$  et  $\varphi_2$  de  $\mathcal{F}$  on a  $\langle T, \ell_1\varphi_1 + \ell_2\varphi_2 \rangle = \ell_1 \langle T, \varphi_1 \rangle + \ell_2 \langle T, \varphi_2 \rangle$ .

L'ensemble des fonctionnelles linéaires et continues sur  $\mathcal{F}$ , noté  $\mathcal{F}'$ , est appelé le dual d'un espace fonctionnel topologique  $\mathcal{F}$ .

**Définition III.5 (Fonction test, espace des fonctions test ([9] page 88))** Soit  $\varphi$  une fonction de  $L(\Omega)$  infiniment dérivable et à support compact<sup>1</sup>. La fonction  $\varphi$  est appelée fonction test ou fonction d'essai.

L'ensemble des fonctions  $\varphi$ , noté  $\mathcal{D}_\infty(\Omega)$ , est appelé espace des fonctions test [98].

Par abus de notation, on note alors  $\mathcal{D}_m(\Omega)$  l'espace des fonctions de  $L(\Omega)$  qui sont  $m$  fois dérivables et à support compact.

**Définition III.6 (Distribution ([9] page 88))** On appelle distribution toute fonctionnelle linéaire et continue sur  $\mathcal{D}_\infty(\Omega)$ .

L'ensemble des distributions est noté  $\mathcal{D}'_\infty(\Omega)$ .

**Définition III.7 (Distribution régulière et distribution singulière ([20] page 35))**

Soit  $\lambda$  la mesure de Lebesgue sur  $\mathbb{R}$  et soit  $s \in L^1_{loc}(\Omega)$ , l'application de  $\mathcal{D}_\infty(\Omega)$  à valeurs dans  $\mathbb{R}$  définie par :

$$T_s : \varphi \mapsto \langle T_s, \varphi \rangle = \int_{\Omega} s(u)\varphi(u)d\lambda(u),$$

est une distribution dite régulière. Toute distribution qui ne peut pas s'écrire sous cette forme est dite singulière.

Pour les distributions régulières, nous dirons par abus de langage que  $s$  est une distribution en pensant en fait à  $T_s$ .

**Exemple III.8 (Distribution de Heaviside et distribution de Dirac ([74] pages 27-28))**

---

1. Une fonction  $\varphi$  est appelée fonction à support compact si  $\text{supp}(\varphi)$  est borné, où  $\text{supp}(\cdot)$  est le support de fonction.

Les deux exemples les plus connus, et les plus utilisés, des distributions sont respectivement la distribution de Heaviside, utilisée pour définir les signaux causaux<sup>2</sup>, et la distribution de Dirac, qui permet de modéliser l'échantillonnage. Pour ce qui est de la distribution de Heaviside, elle peut être facilement définie à partir de la fonction de Heaviside  $H \in L^1_{loc}(\Omega)$ , définie par (I.40). Cette distribution est régulière, notée  $H$  au lieu de  $T_H$ , et définie par :

$$\forall \varphi \in \mathcal{D}_\infty(\Omega), \quad \langle H, \varphi \rangle = \int_\Omega H(u)\varphi(u)d\lambda(u) = \int_0^\infty \varphi(u)du. \quad (\text{III.8})$$

La distribution de Dirac, notée  $\delta \in \mathcal{D}'_\infty(\Omega)$ , est définie par :  $\forall \varphi \in \mathcal{D}_\infty(\Omega)$ ,  $\langle \delta, \varphi \rangle = \varphi(0)$ . Cette distribution est singulière car il n'existe pas de fonction  $\delta$  intégrable telle que :

$$\forall \varphi \in \mathcal{D}_\infty(\Omega), \quad \int_\Omega \delta(u)\varphi(u)d\lambda(u) = \varphi(0).$$

**Définition III.9 (Dérivée d'une distribution)** Soit  $T \in \mathcal{D}'_\infty(\Omega)$ , on appelle dérivée de  $T$ , notée  $dT$ , la distribution définie par :

$$\forall \varphi \in \mathcal{D}_1(\Omega), \quad \langle dT, \varphi \rangle = \langle T, d\varphi \rangle. \quad (\text{III.9})$$

**Propriété III.10 (Dérivée au sens des distributions ([20] page 38))** Pour tout  $m \in \mathbb{N}$  et toute distribution  $T \in \mathcal{D}'_\infty(\Omega)$ , on a :

$$\forall \varphi \in \mathcal{D}_m(\Omega), \quad \langle d^{(i)}T, \varphi \rangle = (-1)^i \langle T, d^{(i)}\varphi \rangle, \forall 0 \leq i \leq m, \quad (\text{III.10})$$

où  $d^{(i)}$  est la dérivée d'ordre  $i$ .

Le cas particulier qui nous intéresse ici est  $m = 1$ . Dans ce cas :

$$\forall T \in \mathcal{D}'_\infty(\Omega), \forall \varphi \in \mathcal{D}_1(\Omega), \quad \langle dT, \varphi \rangle = - \langle T, d\varphi \rangle. \quad (\text{III.11})$$

**Remarque :** Si  $T_s$  est une distribution régulière définie par une fonction  $s$  intégrable sur  $\Omega$ , une dérivation par partie montre que la dérivée d'ordre  $i$  de  $T_s$ , au sens des distributions, coïncide avec la dérivée d'ordre  $i$  de  $s$  au sens des fonctions. En particulier, si  $T_s$  est une distribution régulière associée à  $s \in L^1_{loc}(\Omega)$ , on a :

$$dT_s = T_{ds}. \quad (\text{III.12})$$

◇

On peut remarquer facilement que la dérivée de la distribution de Heaviside est la distribution de Dirac. En effet, si  $H$  est la distribution de Heaviside alors,  $\forall \varphi \in \mathcal{D}_1(\Omega)$ , on a :

$$\begin{aligned} \langle dH, \varphi \rangle &= - \langle H, d\varphi \rangle \\ &= - \int_0^\infty d\varphi(u)du \\ &= - \left[ \varphi(u) \right]_0^\infty = \varphi(0) - \varphi(\infty) \text{ (or } \varphi(\infty) = 0, \text{ car } \varphi \text{ est de support compact)} \\ &= \varphi(0) \\ &= \langle \delta, \varphi \rangle. \end{aligned}$$

---

2. Un signal  $s$  est dit causal, si  $s(t) = 0, \forall t < 0$ .

On déduit de cette propriété le fait que la mesure empirique  $e_n$  est la dérivée, au sens des distributions, de la fonction de répartition empirique  $E_n$ .

Dans la suite de ce manuscrit, nous appelons  $\mathcal{K}_1(\Omega)$  l'ensemble des noyaux sommatifs de  $\mathcal{K}(\Omega)$  qui sont au moins une fois dérivable et à support compact. Notons que  $\mathcal{K}_1(\Omega) \subset \mathcal{D}_1(\Omega)$ . Donc par définition, tout noyau sommatif une fois différentiable vérifie l'expression (III.11).

On peut alors reformuler l'estimateur de densité de Parzen-Rosenblatt  $\widehat{f}_\kappa^n$ , en chaque point  $x \in \Omega$ , comme le produit scalaire de la fonction de répartition empirique  $E_n$  avec la dérivée d'un noyau sommatif  $\kappa$  translaté en  $x$  :

$$\widehat{f}_\kappa^n(x) = \langle E_n, -d\kappa^x \rangle. \quad (\text{III.13})$$

**Preuve :**  $\forall x \in \Omega$ , on a :

$$\begin{aligned} \widehat{f}_\kappa^n(x) &= \langle e_n, \kappa^x \rangle \\ &= \langle dE_n, \kappa^x \rangle \text{ par l'expression (III.11)} \\ &= \langle E_n, -d\kappa^x \rangle. \end{aligned}$$

□

### III.1.3 Décomposition de la dérivée des noyaux sommatifs

Nous ne considérons, dans ce travail, que les noyaux sommatifs de  $\mathcal{K}_1(\Omega)$ , c'est-à-dire qui sont monomodaux, symétriques, centrés avec un support borné et sont au moins une fois dérivable. Soit  $\kappa \in \mathcal{K}_1(\Omega)$ . On définit la fonction  $d\kappa^+$  de  $\Omega$  à valeurs dans  $\mathbb{R}_+$  par :  $\forall u \in \Omega, d\kappa^+(u) = \max(0, -d\kappa(u))$ . On définit également  $d\kappa^-(u) = \max(0, d\kappa(u))$ . Par construction,  $-d\kappa = d\kappa^+ - d\kappa^-$ . Cette décomposition est appelée la **décomposition de Jordan** de la dérivée première d'un noyau sommatif  $\kappa$ . La reformulation pratique de l'estimateur de densité de Parzen-Rosenblatt sur laquelle nous appuyons notre travail est essentiellement basée sur cette décomposition.

La dérivée d'un noyau sommatif peut alors être décomposée via la propriété suivante :

#### Propriété III.11 (Décomposition sommative de la dérivée du noyau sommatif)

La dérivée d'un noyau sommatif  $\kappa \in \mathcal{K}_1(\Omega)$  peut s'écrire comme la combinaison linéaire de deux noyaux sommatifs  $\eta^+$  et  $\eta^-$  :

$$\forall u \in \Omega, -d\kappa(u) = a \left( \eta^+(u) - \eta^-(u) \right), \quad (\text{III.14})$$

où  $a = \int_\Omega \max(0, -d\kappa(u)) du$ ,  $\eta^+(u) = \frac{d\kappa^+(u)}{a}$  et  $\eta^-(u) = \frac{d\kappa^-(u)}{a}$ .

**Preuve :** Par construction,  $\eta^+$  et  $\eta^-$  sont positifs et sommatifs. En effet,

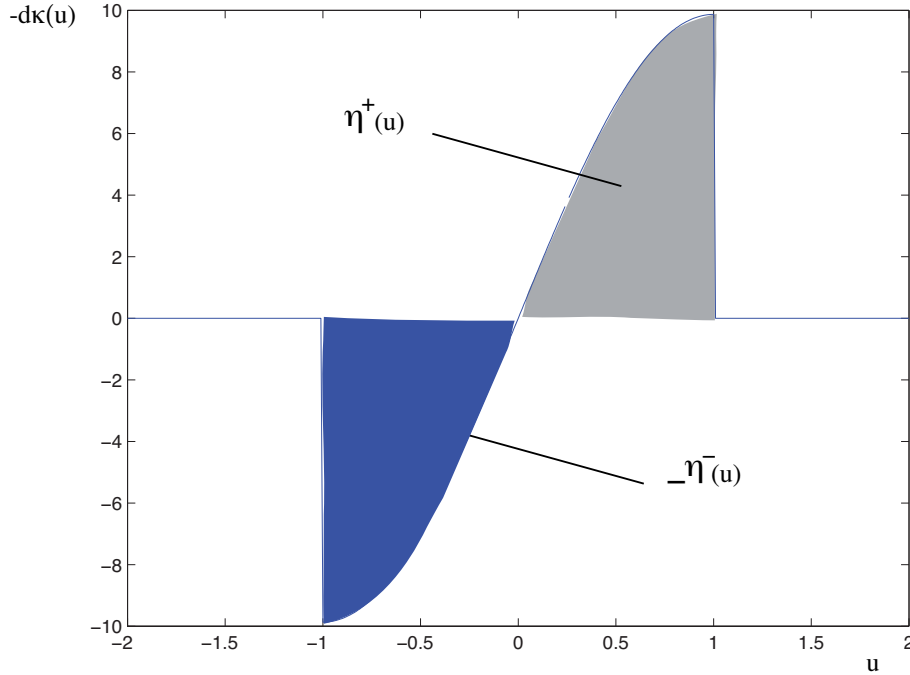
$$\int_\Omega \eta^+(u) du = \frac{1}{a} \int_\Omega d\kappa^+(u) du = 1,$$

et

$$\int_\Omega \eta^-(u) du = \frac{1}{a} \int_\Omega d\kappa^-(u) du = 1.$$

Par construction,  $d\kappa^+ = a \eta^+$  et  $d\kappa^- = a \eta^-$ , d'où l'expression (III.14). □

La Figure III.1 illustre la décomposition sommative de la dérivée d'un noyau sommatif  $\kappa$ . Sur la Figure III.1 le noyau sommatif  $\eta^+$  est dessiné en gris et le noyau  $\eta^-$  en bleu.



**Figure III.1** : La dérivée d'un noyau sommatif  $\kappa$  comme la combinaison linéaire de deux noyaux sommatifs  $\eta^+$  et  $\eta^-$ .

La plupart des noyaux sommatifs sont définis à partir d'une fonction générique  $\kappa$  dilatée par un facteur positif non nul  $\Delta$ , par  $\forall u \in \Omega$ ,  $\kappa_\Delta(u) = \frac{1}{\Delta} \kappa(\frac{u}{\Delta})$ . Alors  $a_\Delta$ , la constante associée à la décomposition sommative de la dérivée première d'un noyau sommatif  $\kappa_\Delta$ , peut facilement être déduite de  $a = a_1$  par  $a_\Delta = \frac{a}{\Delta}$ .

**Propriété III.12 (Décomposition sommative de la dérivée d'un noyau dilaté)**  
Soit  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première est  $(a, \eta^+, \eta^-)$  et soit  $\Delta$  un réel positif non nul, alors la décomposition sommative de la dérivée première de  $\kappa_\Delta$  est  $(a_\Delta, \eta_\Delta^+, \eta_\Delta^-)$ , avec  $a_\Delta = \frac{a}{\Delta}$ .

**Preuve** :  $\forall u \in \Omega$ , on a :

$$d\kappa_\Delta(u) = \frac{1}{\Delta^2} d\kappa\left(\frac{u}{\Delta}\right). \quad (\text{III.15})$$

En effet,

$\kappa_\Delta(u) = \frac{1}{\Delta} \kappa\left(\frac{u}{\Delta}\right)$ , par le changement de variable  $\frac{u}{\Delta} = v$ , la dérivée du noyau sommatif  $\kappa_\Delta$ , s'exprime par (III.15).

Ensuite, d'après l'expression (III.14), on a :

$$\begin{aligned} d\kappa_{\Delta}(u) &= \frac{a}{\Delta^2} \left( \eta^+\left(\frac{u}{\Delta}\right) - \eta^-\left(\frac{u}{\Delta}\right) \right) \\ &= \frac{a}{\Delta} \left( \frac{1}{\Delta} \eta^+\left(\frac{u}{\Delta}\right) - \frac{1}{\Delta} \eta^-\left(\frac{u}{\Delta}\right) \right) \\ &= a_{\Delta} \left( \eta_{\Delta}^+(u) - \eta_{\Delta}^-(u) \right). \end{aligned}$$

□

### III.1.4 Dérivée de noyau et estimateur de densité de Parzen-Rosenblatt

#### III.1.4.1 Estimateur de densité de Parzen-Rosenblatt

Soient  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première est  $(a, \eta^+, \eta^-)$ . La dérivée du noyau sommatif  $\kappa$  translaté en  $x \in \Omega$  est donnée par :

$$\forall u \in \Omega, d\kappa^x(u) = d\kappa(u - x). \quad (\text{III.16})$$

La décomposition de Jordan d'un noyau sommatif translaté peut alors être donnée via la propriété suivante :

**Propriété III.13 (Décomposition de Jordan d'un noyau sommatif translaté)** *Soient  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition de Jordan de la dérivée première est  $(d\kappa^+, d\kappa^-)$  et  $x \in \Omega$ , alors la décomposition de Jordan de la dérivée première de  $\kappa^x$  est  $(d\kappa^{x+}, d\kappa^{x-})$ , avec  $d\kappa^{x+} = \max(0, -d\kappa^x)$  et  $d\kappa^{x-} = \max(0, d\kappa^x)$ .*

**Preuve :** D'après l'expression (III.16),  $\forall u \in \Omega$ , on a :

$$\begin{aligned} d\kappa^{x+}(u) &= d\kappa^+(u - x) \\ &= \max\left(0, -d\kappa(u - x)\right) \\ &= \max\left(0, -d\kappa^x(u)\right), \end{aligned}$$

et

$$\begin{aligned} d\kappa^{x-}(u) &= d\kappa^-(u - x) \\ &= \max\left(0, d\kappa(u - x)\right) \\ &= \max\left(0, d\kappa^x(u)\right). \end{aligned}$$

□

**Propriété III.14 (Décomposition sommative de la dérivée d'un noyau translaté)**

Soient  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première est  $(a, \eta^+, \eta^-)$  et  $x \in \Omega$ , alors la décomposition sommative de la dérivée première de  $\kappa^x$  est  $(a, \eta^{x+}, \eta^{x-})$ , avec  $\eta^{x+} = \frac{d\kappa^{x+}}{a}$  et  $\eta^{x-} = \frac{d\kappa^{x-}}{a}$ .

**Preuve :** Soit  $x \in \Omega$ . Par construction,  $\eta^{x+}$  et  $\eta^{x-}$  sont positifs et sommatifs. En effet,

$$\int_{\Omega} \eta^{x+}(u) du = \int_{\Omega} \eta^+(u-x) du = \int_{\Omega} \eta^+(u) du = 1,$$

et

$$\int_{\Omega} \eta^{x-}(u) du = \int_{\Omega} \eta^-(u-x) du = \int_{\Omega} \eta^-(u) du = 1.$$

Par construction,  $d\kappa^{x+} = a \eta^{x+}$  et  $d\kappa^{x-} = a \eta^{x-}$ , d'où  $-d\kappa^x = a(\eta^{x+} - \eta^{x-})$ . □

Le théorème suivant, montre que l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_{\kappa}^n$  en chaque point  $x \in \Omega$ , peut être vu comme la combinaison linéaire de deux estimateurs  $\hat{F}_{\eta^+}^n$  (rsp.  $\hat{F}_{\eta^-}^n$ ) à noyau  $\eta^+$  (rsp.  $\eta^-$ ) de Parzen-Rosenblatt de la fonction de répartition en chaque point  $x \in \Omega$ .

**Théorème III.15 (Première reformulation de l'estimateur de Parzen-Rosenblatt)**

Soit  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première est  $(a, \eta^+, \eta^-)$ , pour tout  $x \in \Omega$ , on a :

$$\hat{f}_{\kappa}^n(x) = \frac{a}{\Delta} \left( \hat{F}_{\eta^+}^n(x) - \hat{F}_{\eta^-}^n(x) \right), \quad (\text{III.17})$$

où  $\hat{F}_{\xi}^n$  est l'estimateur à noyau  $\xi$  de Parzen-Rosenblatt de la fonction de répartition.

**Preuve :**  $\forall x \in \Omega$ , on a :

$$\begin{aligned} \hat{f}_{\kappa}^n(x) &= \langle E_n, -d\kappa^x \rangle \text{ par la propriété III.14} \\ &= \langle E_n, a(\eta^{x+} - \eta^{x-}) \rangle \text{ par linéarité du produit scalaire} \\ &= a \left( \langle E_n, \eta^{x+} \rangle - \langle E_n, \eta^{x-} \rangle \right) \\ &= a(\mathbb{E}_{\eta^{x+}}(E_n) - \mathbb{E}_{\eta^{x-}}(E_n)) \text{ par l'expression (II.79)} \\ &= a \left( \hat{F}_{\eta^+}^n(x) - \hat{F}_{\eta^-}^n(x) \right). \end{aligned}$$

□

### III.1.4.2 Estimateur de densité de Parzen-Rosenblatt : les deux noyaux de la décomposition sont symétriques et issus de la translation d'un même noyau sommatif

Certains noyaux de  $\mathcal{K}_1(\Omega)$  sont tels que les deux noyaux  $\eta^+$  et  $\eta^-$  dans la décomposition sont les translatés d'un même noyau sommatif symétrique  $\eta$ . C'est le cas par exemple du noyau sommatif défini par :

$$\kappa(u) = \frac{1}{2} \left( 1 + \cos(|u|\pi) \right) \mathbb{1}_{[-1,1]}(u),$$

où les deux noyaux sommatifs de décomposition sont données par :

$$\eta^+(u) = \frac{\pi}{2} \sin(u\pi) \mathbb{1}_{[0,1]}(u) \quad \text{et} \quad \eta^-(u) = \frac{\pi}{2} \sin(u\pi) \mathbb{1}_{[-1,0]}(u).$$

Ces deux noyaux sommatifs sont les translatés d'un noyau sommatif symétrique  $\eta$  défini par :

$$\eta(u) = \frac{\pi}{2} \left( \cos(|u|\pi) \right) \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u).$$

Par construction cette translation est liée à la largeur de bande du noyau. On a donc :

$$\forall u \in \Omega, \quad \eta^+(u) = \eta\left(u - \frac{1}{2}\right) \quad \text{et} \quad \eta^-(u) = \eta\left(u + \frac{1}{2}\right). \quad (\text{III.18})$$

**Remarque** : Par construction, pour tout  $\Delta \in \mathbb{R}_+^*$  et  $u \in \Omega$ , on a :

$$\eta_\Delta^+(u) = \frac{1}{\Delta} \eta\left(\frac{u}{\Delta} - \frac{1}{2\Delta}\right) = \eta_\Delta\left(u - \frac{\Delta}{2}\right),$$

et

$$\eta_\Delta^-(u) = \frac{1}{\Delta} \eta\left(\frac{u}{\Delta} + \frac{1}{2\Delta}\right) = \eta_\Delta\left(u + \frac{\Delta}{2}\right).$$

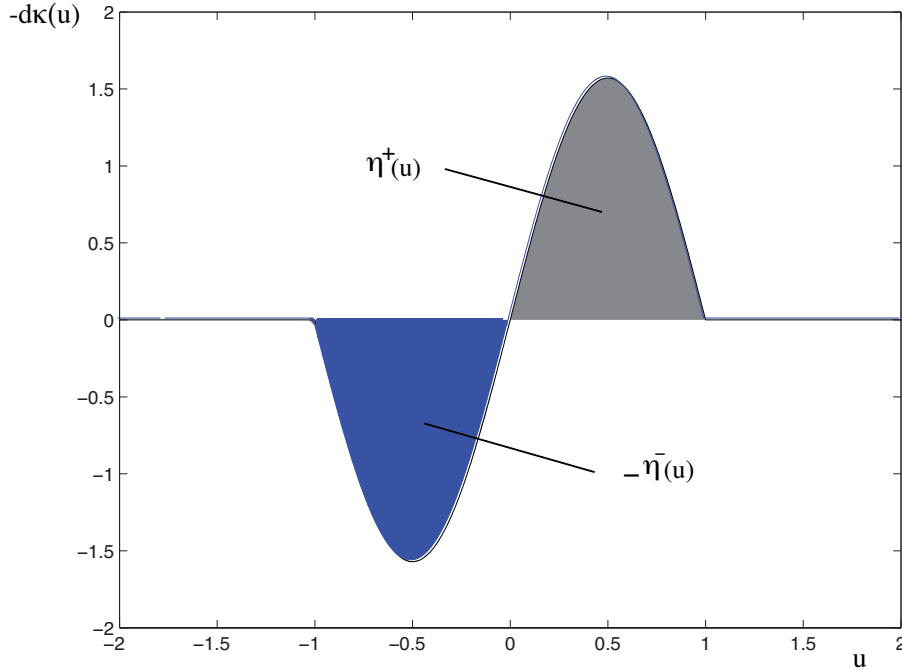
◇

Dans ce cas, la dérivée du noyau  $\kappa$  s'écrit  $d\kappa$  et vérifie :

$$\forall u \in \Omega, \quad -d\kappa(u) = a \left( \eta\left(u - \frac{1}{2}\right) - \eta\left(u + \frac{1}{2}\right) \right), \quad (\text{III.19})$$

où  $a = \int_\Omega \max(0, -d\kappa(u)) du$ .

La Figure III.2 illustre la décomposition sommative de la dérivée première d'un noyau sommatif  $\kappa$  dont les deux noyaux  $\eta^+$  et  $\eta^-$  sont symétriques et translatés d'un même noyau sommatif. Sur la Figure III.2 le noyau sommatif  $\eta^+$  est dessiné en gris et le noyau  $\eta^-$  en bleu.



**Figure III.2 :** Décomposition sommative de la dérivée première de  $\kappa$  : cas où les deux noyaux  $\eta^+$  et  $\eta^-$  sont symétriques et issus de la translation d'un même noyau sommatif.

Dû à cette simplification de la décomposition sommative de la dérivée de  $\kappa$ , l'expression de l'estimateur de Parzen-Rosenblatt prend une autre forme. Remarquons tout d'abord que, pour tout  $x \in \Omega$ , les noyaux translétés  $\eta^{x+}$  et  $\eta^{x-}$  se réécrivent :

$$\forall u \in \Omega, \eta^{x+}(u) = \eta(u - x - \frac{1}{2}) \text{ et } \eta^{x-}(u) = \eta(u - x + \frac{1}{2}). \quad (\text{III.20})$$

Et donc la dérivée de  $\kappa^x$  se décompose en :

$$\forall u \in \Omega, -d\kappa^x(u) = a \left( \eta(u - x - \frac{1}{2}) - \eta(u - x + \frac{1}{2}) \right). \quad (\text{III.21})$$

On définit alors  $E_n^+$  et  $E_n^-$ , deux fonctions de répartition empirique, par :

$$\forall u \in \Omega, E_n^+(u) = E_n(u + \frac{1}{2}) \text{ et } E_n^-(u) = E_n(u - \frac{1}{2}), \quad (\text{III.22})$$

où  $E_n$  est la fonction de répartition empirique définie par (I.39).

Le théorème suivant montre que l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_\kappa^n$ , en chaque point  $x \in \Omega$ , peut être vu comme l'espérance mathématique de la combinaison linéaire des deux fonctions de répartition empirique  $E_n^+$  et  $E_n^-$  dans un voisinage probabiliste défini par le noyau sommatif  $\eta$  translété en  $x$ .

**Théorème III.16 (Deuxième reformulation de l'estimateur de Parzen-Rosenblatt)**

Soit  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première s'écrit, pour tout  $u \in \Omega$  :  $-d\kappa(u) = a \left( \eta(u - \frac{1}{2}) - \eta(u + \frac{1}{2}) \right)$ , avec  $a = \int_\Omega \max(0, -d\kappa(u)) du$ . Pour tout  $x \in \Omega$ , on a :

$$\hat{f}_\kappa^n(x) = a \mathbb{E}_{\eta^x} (E_n^+ - E_n^-), \quad (\text{III.23})$$



où  $E_n^+$  et  $E_n^-$  sont deux fonctions de répartition empirique définies dans (III.22) .

**Preuve :** D'après l'expression (III.13),  $\forall x \in \Omega$ , on a :

$$\begin{aligned}
\widehat{f}_\kappa^n(x) &= \langle E_n, -d\kappa^x \rangle \text{ par l'expression (III.3)} \\
&= \int_\Omega -d\kappa(u-x)E_n(u)du \text{ par l'expression (III.21)} \\
&= a \left( \int_\Omega \eta(u-x-\frac{1}{2})E_n(u)du - \int_\Omega \eta(u-x+\frac{1}{2})E_n(u)du \right) \\
&= a \left( \int_\Omega \eta(u-x)E_n(u+\frac{1}{2})du - \int_\Omega \eta(u-x)E_n(u-\frac{1}{2})du \right) \\
&= a \left( \int_\Omega \eta(u-x) (E_n^+(u) - E_n^-(u)) du \right) \\
&= a \mathbb{E}_{\eta^x} (E_n^+ - E_n^-) .
\end{aligned}$$

La quatrième ligne de preuve est obtenue par le changement de variable  $u + \frac{1}{2} = v$  et  $u - \frac{1}{2} = v$ . □

### III.1.4.3 Estimateur de densité de Parzen-Rosenblatt : cas où le noyau est simplement symétrique

Soit  $\kappa \in \mathcal{K}_1(\Omega)$ , le noyau sommatif utilisé pour estimer l'estimateur de densité de Parzen-Rosenblatt  $\widehat{f}_\kappa^n$  en chaque point  $x \in \Omega$ .  $\kappa$  étant symétrique, la décomposition de sa dérivée première peut être exprimée avec un seul noyau sommatif  $\eta$ . Par construction, les deux noyaux  $\eta^+$  et  $\eta^-$  sont donnés par :

$$\forall u \in \Omega, \eta^+(u) = \eta\left(\frac{1}{2} - u\right) \text{ et } \eta^-(u) = \eta\left(\frac{1}{2} + u\right). \quad (\text{III.24})$$

**Remarque :** Par construction, pour tout  $\Delta \in \mathbb{R}_+^*$  et  $u \in \Omega$ , on a :

$$\eta_\Delta^+(u) = \frac{1}{\Delta} \eta\left(\frac{1}{2\Delta} - \frac{u}{\Delta}\right) = \eta_\Delta\left(\frac{\Delta}{2} - u\right),$$

et

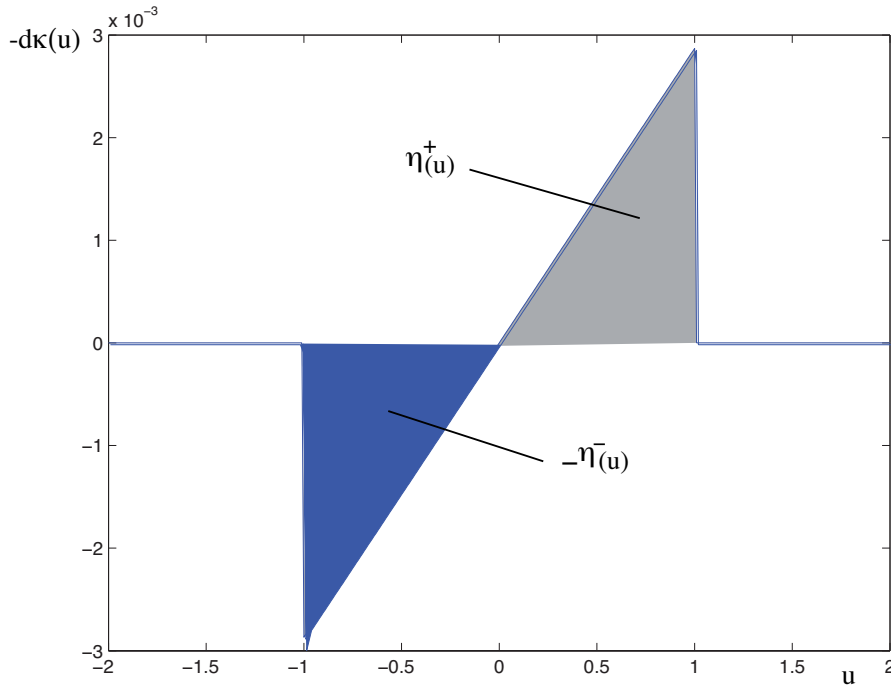
$$\eta_\Delta^-(u) = \frac{1}{\Delta} \eta\left(\frac{1}{2\Delta} + \frac{u}{\Delta}\right) = \eta_\Delta\left(\frac{\Delta}{2} + u\right).$$

◇

Dans ce cas, la dérivée du noyau  $\kappa$  s'écrit  $d\kappa$  et vérifie :

$$\forall u \in \Omega, -d\kappa(u) = a \left( \eta\left(\frac{1}{2} - u\right) - \eta\left(\frac{1}{2} + u\right) \right). \quad (\text{III.25})$$

où  $a = \int_\Omega \max(0, -d\kappa(u))du$ . La Figure III.3 illustre la décomposition sommatif de la dérivée première d'un noyau sommatif symétrique  $\kappa$ . Sur la Figure III.3 le noyau sommatif  $\eta^+$  est dessinée en gris et le noyau  $\eta^-$  en bleu.



**Figure III.3 :** Décomposition sommative de la dérivée première de  $\kappa$  : cas où le noyau sommatif  $\kappa$  est simplement symétrique.

Cette simplification de la décomposition sommative de la dérivée de  $\kappa$ , nous conduit à une autre reformulation de l'estimateur de Parzen-Rosenblatt. Remarquons tout d'abord que, pour tout  $x \in \Omega$ , les noyaux tradlatés  $\eta^{x+}$  et  $\eta^{x-}$  se réécrivent :

$$\forall u \in \Omega, \eta^{x+}(u) = \eta\left(\frac{1}{2} - u - x\right) \text{ et } \eta^{x-}(u) = \eta\left(\frac{1}{2} + u - x\right). \quad (\text{III.26})$$

Et donc la dérivée de  $\kappa^x$  se décompose en :

$$\forall u \in \Omega, -d\kappa^x(u) = a \left( \eta\left(\frac{1}{2} - u - x\right) - \eta\left(\frac{1}{2} + u - x\right) \right). \quad (\text{III.27})$$

On définit alors  $E_n^{x+}$  et  $E_n^{x-}$ , deux fonctions de répartition empirique tradlatés en  $x$ , par :

$$\forall u \in \Omega, E_n^{x+}(u) = E_n\left(x + \frac{1}{2} - u\right) \text{ et } E_n^{x-}(u) = E_n\left(x - \frac{1}{2} + u\right). \quad (\text{III.28})$$

L'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_\kappa^n$ , en chaque point  $x \in \Omega$ , peut être vu comme l'espérance mathématique de la combinaison linéaire de deux fonctions de répartition empirique  $E_n^{x+}$  et  $E_n^{x-}$  dans un voisinage probabiliste défini par le noyau sommatif  $\eta$ .

**Théorème III.17 (Troisième reformulation de l'estimateur de Parzen-Rosenblatt)**

Soit  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première s'écrit, pour tout  $u \in \Omega$  :  $-d\kappa(u) = a\left(\eta\left(\frac{1}{2} - u\right) - \eta\left(\frac{1}{2} + u\right)\right)$ , avec  $a = \int_\Omega \max(0, -d\kappa(u))du$ . Pour tout  $x \in \Omega$ , on a :

$$\hat{f}_\kappa^n(x) = a \mathbb{E}_\eta \left( E_n^{x-} - E_n^{x+} \right), \quad (\text{III.29})$$

où  $E_n^{x-}$  et  $E_n^{x+}$  sont deux fonctions de répartition empirique définies dans (III.28).

**Preuve :** D'après l'expression (III.13),  $\forall x \in \Omega$ , on a :

$$\begin{aligned}
\widehat{f}_\kappa^n(x) &= \langle E_n, -d\kappa^x \rangle, \text{ par l'expression (III.3)} \\
&= \int_{-\infty}^{+\infty} -d\kappa(u-x)E_n(u)du \text{ par l'expression (III.27)} \\
&= a \left( \int_{-\infty}^{+\infty} \eta\left(\frac{1}{2} - u + x\right)E_n(u)du - \int_{-\infty}^{+\infty} \eta\left(\frac{1}{2} + u - x\right)E_n(u)du \right) \\
&= a \left( \int_{-\infty}^{+\infty} \eta(u)E_n\left(x + \frac{1}{2} - u\right)du - \int_{-\infty}^{+\infty} \eta(u)E_n\left(x - \frac{1}{2} + u\right)du \right) \\
&= a \left( \int_{-\infty}^{+\infty} \eta(u) \left( E_n^{x-}(u) - E_n^{x+}(u) \right) du \right) \\
&= a \mathbb{E}_\eta \left( E_n^{x-} - E_n^{x+} \right).
\end{aligned}$$

La partie droite de quatrième ligne de preuve est obtenue par le changement de variable  $\frac{1}{2} + u - x = v$ , tandis que la partie de gauche est obtenue par le changement de variable  $\frac{1}{2} - u + x = v$ .

□

## III.2 Estimation maxitive de la densité

Cette section présente le coeur des travaux de ce mémoire. Nous y proposons une extension de la méthode d'estimation de Parzen-Rosenblatt. Le principe de cette extension est simple. Il s'agit de remplacer l'estimation précise proposée par la théorie classique de l'estimation non-paramétrique par une estimation imprécise. Cette estimation imprécise a pour but de définir l'enveloppe convexe de toutes les estimations précises qui auraient été obtenues pour un ensemble de noyaux ayant des propriétés analogues. Elle s'appuie sur la possibilité de définir des familles de noyaux sommatifs avec des noyaux maxitifs. C'est pourquoi nous l'appelons **estimation maxitive de la densité**. Ce que l'on gagne généralement en étendant une méthode précise d'estimation, en remplaçant les noyaux sommatifs par des noyaux maxitifs, est l'obtention d'une grandeur de sortie de nature intervalliste dont l'imprécision est représentative d'un (ou plusieurs) défaut(s) du processus d'estimation. Par exemple, dans [73], il est montré que, dans certaines conditions, l'imprécision de la sortie d'un processus de filtrage maxitif permet de quantifier l'impact, sur cette même sortie, des variations aléatoires de la grandeur d'entrée. On retrouve à peu près la même propriété dans le cas présent, comme nous allons le montrer dans la partie expérimentale : l'imprécision de l'estimation maxitive de la densité reflète l'amplitude des variations statistiques de l'estimation dues aux variations aléatoires des observations. Elle peut donc permettre une analyse de la sensibilité de l'estimateur. Cette analyse peut être obtenue sans accroître de façon sensible la complexité algorithmique de l'estimation.

### III.2.1 Principe et définition de l'estimation maxitive de la densité

Comme toutes les extensions maxitives, le principe de l'estimation de la densité que nous proposons consiste en le remplacement d'un (ou plusieurs) noyau(x) sommatif(s)

par un (ou plusieurs) noyau(x) maxitif(s) qui le(s) domine(nt). Cette substitution induit l'utilisation de l'extension de l'opérateur espérance présenté en section II.1.2.2.

En section II.3.3, nous avons présenté l'extension maxitive de l'estimateur de Parzen-Rosenblatt de la fonction de répartition. Cette extension, en chaque point  $x$  de  $\Omega$  s'écrit simplement :

$$\forall x \in \Omega, \overline{F}_\pi^n(x) = [\underline{F}_\pi^n(x), \overline{F}_\pi^n(x)] = \overline{\mathbb{E}}_{\pi^x}^n(E_n) = [\underline{\mathbb{E}}_{\pi^x}^n(E_n), \overline{\mathbb{E}}_{\pi^x}^n(E_n)],$$

où  $E_n$  est la fonction de répartition empirique,  $\overline{\mathbb{E}}$  est l'extension de l'opérateur espérance aux capacités concaves et  $\pi$  est un noyau maxitif.

L'extension maxitive de l'estimateur de densité de Parzen-Rosenblatt ne peut être aussi directe à cause des limites imposées par l'utilisation de l'intégrale de Choquet. La construction que nous proposons s'appuie sur la décomposition sommative présentée en section III.1.3. Une telle construction comprend quatre étapes principales.

### Etape 1 : choix d'un noyau sommatif de référence

Cette première étape est une étape d'expertise. L'utilisateur choisit un noyau et une largeur de bande en fonction des connaissances qu'il a du processus d'observation et/ou d'une analyse préliminaire des données<sup>3</sup>. Dans le cas présent, le noyau  $\kappa$  de référence doit être choisi dans la famille  $\mathcal{K}_1(\Omega)$  définie précédemment.

### Etape 2 : décomposition sommative de la dérivée première du noyau de référence

Cette décomposition aboutit au triplet  $(a, \eta^+, \eta^-)$  avec  $a \in \mathbb{R}^+$ ,  $\eta^+ \in \mathcal{K}(\Omega)$  et  $\eta^- \in \mathcal{K}(\Omega)$ .

### Etape 3 : domination de deux noyaux sommatifs de la décomposition sommative.

Cette troisième étape consiste à définir deux noyaux maxitifs  $\pi^+$  et  $\pi^-$  dominant les deux noyaux sommatifs  $\eta^+$  et  $\eta^-$ . Les techniques de domination sont présentées dans ([69], pages 73-75). Pour assurer une spécificité maximale à l'estimation imprécise de la densité, une méthode pratique consiste à utiliser les deux noyaux maxitifs les plus spécifiques dominant les deux noyaux sommatifs. On utilise pour cela la transformation probabilité possibilité [37, 38]. Cette domination permet de définir un ensemble des noyaux sommatifs que nous notons  $\mathcal{D}(a, (\pi^+, \pi^-))$  par :

$$\mathcal{D}(a, (\pi^+, \pi^-)) = \left\{ v \in \mathcal{K}_1(\Omega), \exists \xi^+ \in \mathcal{M}(\pi^+) \text{ et } \xi^- \in \mathcal{M}(\pi^-) \right. \\ \left. \text{tel que } -dv = a(\xi^+ - \xi^-) \right\}, \quad (\text{III.30})$$

où  $\mathcal{M}(\pi^+)$  (rsp.  $\mathcal{M}(\pi^-)$ ) est la famille des noyaux sommatifs dominés par le noyau maxitif  $\pi^+$  (rsp.  $\pi^-$ ).

Cet ensemble est naturellement non vide car, par construction,  $\kappa \in \mathcal{D}(a, (\pi^+, \pi^-))$ .

---

3. Un tel choix d'expert peut par exemple comporter une première phase d'analyse grossière des données basée sur un histogramme, lui permettant de spécifier la largeur de bande du noyau. Le choix du noyau d'Epanechnikov comme fonction associée au noyau peut être motivée par les résultats sur le comportement asymptotique des estimateurs de Parzen-Rosenblatt montrant que ce noyau est celui qui minimise la distance entre l'estimateur et la densité (le AMISE) [106, 100].

**Etape 4 : estimation maxitive de la densité.**

Dans cette ultime étape, on construit l'estimation maxitive de la densité proprement dite, notée  $\underline{f}_{(\kappa)}^n$ , en utilisant les noyaux maxitifs translatsés  $\pi^+$  et  $\pi^-$  par :

$$\forall x \in \Omega, \underline{f}_{(\kappa)}^n(x) = a \left( \underline{\mathbb{E}}_{\pi^{x+}}(E_n) \ominus \underline{\mathbb{E}}_{\pi^{x-}}(E_n) \right), \quad (\text{III.31})$$

où  $\ominus$  est la différence de Minkowski [17] et  $\underline{\mathbb{E}}_{\pi^+}(\cdot)$  (rsp.  $\underline{\mathbb{E}}_{\pi^-}(\cdot)$ ) est l'espérance imprécise basée sur le noyau maxitif  $\pi^+$  (rsp.  $\pi^-$ ).

On peut facilement développer l'expression (III.31) pour faire apparaître des estimations maxitives de la fonction de répartition empirique. Pour tout  $x \in \Omega$ , on a :

$$\begin{aligned} \underline{f}_{(\kappa)}^n(x) &= \left[ \underline{f}_{(\kappa)}^n(x), \overline{f}_{(\kappa)}^n(x) \right] = a \left( \underline{F}_{\pi^+}^n(x) \ominus \overline{F}_{\pi^-}^n(x) \right) \\ &= a \left[ \underline{F}_{\pi^+}^n(x) - \overline{F}_{\pi^-}^n(x), \overline{F}_{\pi^+}^n(x) - \underline{F}_{\pi^-}^n(x) \right], \end{aligned} \quad (\text{III.32})$$

où  $\underline{F}_{\pi^+}^n = [\underline{F}_{\pi^+}^n, \overline{F}_{\pi^+}^n]$  et  $\overline{F}_{\pi^-}^n = [\underline{F}_{\pi^-}^n, \overline{F}_{\pi^-}^n]$ .

**III.2.2 Algorithme de calcul de bornes de l'estimation maxitive de la densité**

Lorsqu'il s'agit de représenter graphiquement la densité pour réaliser une analyse experte des observations, ou encore en vue d'une utilisation algorithmique, il est souvent nécessaire d'estimer cette densité sur un ensemble (régulièrement espacés ou non) de points de  $\Omega$ . Soient  $\{y_j\}_{j \in \{1, \dots, p\}}$  ces  $p$  points. L'algorithme de calcul de  $\{\underline{f}_{(\kappa)}^n(y_j)\}_{j \in \{1, \dots, p\}}$ , en chaque point  $\{y_j\}_{j \in \{1, \dots, p\}}$  de  $\Omega$ , découle facilement de l'algorithme de calcul de  $\{\underline{F}_{\pi^+}^n(y_j)\}_{j \in \{1, \dots, p\}}$ . Il s'agit simplement, d'après l'expression (III.32), de calculer les deux intervalles  $\{\underline{F}_{\pi^+}^n(y_j)\}_{j \in \{1, \dots, p\}}$  et  $\{\overline{F}_{\pi^-}^n(y_j)\}_{j \in \{1, \dots, p\}}$ . Le résultat obtenu est multiplié par la constante  $a$ . L'algorithme suivant décrit plus formellement le processus de calcul.

**Data:** les observations  $\{y_j\}_{j \in \{1, \dots, p\}}$ , les noyaux maxitifs  $\pi^+$  et  $\pi^-$  et le facteur  $a$ .

**Result:**  $\left\{ \left[ \underline{f}_{(\kappa)}^n(y_j), \overline{f}_{(\kappa)}^n(y_j) \right] \right\}_{j \in \{1, \dots, p\}}$ .

**begin**

**for**  $j = 1$  *to*  $p$  **do**

    Calculer  $\underline{F}_{\pi^-}^n(y_j)$  et  $\overline{F}_{\pi^+}^n(y_j)$ , par l'expression (II.81).

    Calculer  $\overline{F}_{\pi^+}^n(y_j)$  et  $\underline{F}_{\pi^-}^n(y_j)$ , par l'expression (II.82).

$\overline{f}_{(\kappa)}^n(y_j) = a \left( \overline{F}_{\pi^+}^n(y_j) - \underline{F}_{\pi^-}^n(y_j) \right)$ .

$\underline{f}_{(\kappa)}^n(y_j) = a \left( \underline{F}_{\pi^+}^n(y_j) - \overline{F}_{\pi^-}^n(y_j) \right)$ .

**end**

**end**

**Algorithm 2:** Calcul de  $\left\{ \left[ \underline{f}_{(\kappa)}^n(y_j), \overline{f}_{(\kappa)}^n(y_j) \right] \right\}_{j \in \{1, \dots, p\}}$ .

### III.2.3 Relation entre estimation maxitive et estimation sommative de la densité

Comme nous l'avons dit au paragraphe II.3.3.2, l'estimation maxitive de la fonction de répartition  $\overline{F}_\pi^n$ , est entièrement cohérente avec l'estimation sommative car elle respecte une *double inclusion* :

- première inclusion :  $\forall x \in \Omega, \forall \kappa \in \mathcal{M}(\pi), \widehat{F}_\kappa^n(x) \in \overline{F}_\pi^n(x)$ ,
- deuxième inclusion :  $\forall x \in \Omega, \forall y \in \overline{F}_\pi^n(x), \exists \kappa \in \mathcal{M}(\pi), \widehat{F}_\kappa^n(x) = y$ .

Dans le cas de l'estimation maxitive de la densité  $\underline{f}_{(\kappa)}^n(x)$ , seule la cohérence au sens de la première inclusion est conservée.

#### **Théorème III.18 (Inclusion de l'estimation sommative dans l'extension maxitive)**

Soit  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première est  $(a, \eta^+, \eta^-)$ . Soit  $\pi^+$  (resp.  $\pi^-$ ) un noyau maxitif dominant  $\eta^+$  (resp.  $\eta^-$ ). Pour tout  $x \in \Omega$ , on a :

$$\forall \varphi \in \mathcal{D}(a, (\pi^+, \pi^-)), \widehat{f}_\varphi^n(x) \in \underline{f}_{(\kappa)}^n(x), \quad (\text{III.33})$$

où  $\mathcal{D}(a, (\pi^+, \pi^-))$  est la famille des noyaux sommatifs définie par (III.30) et  $\underline{f}_{(\kappa)}^n$  est l'intervalle d'estimation de la densité défini par (III.31).

**Preuve :** Soit  $\varphi \in \mathcal{D}(a, (\pi^+, \pi^-))$ . Il existe  $\xi^+ \in \mathcal{M}(\pi^+)$  et  $\xi^- \in \mathcal{M}(\pi^-)$ , tels que

$$\forall u \in \Omega, -d\varphi(u) = a(\xi^+(u) - \xi^-(u)).$$

Or, d'après l'expression (II.83), pour tout  $x \in \Omega$ ,

$$\forall \xi^+ \in \mathcal{M}(\pi^+), \widehat{F}_{\xi^+}^n(x) \in \overline{\mathbb{E}}_{\pi^{x^+}}(E_n),$$

et

$$\forall \xi^- \in \mathcal{M}(\pi^-), \widehat{F}_{\xi^-}^n(x) \in \overline{\mathbb{E}}_{\pi^{x^-}}(E_n).$$

D'après l'expression (III.17), on a :

$$\forall x \in \Omega, \widehat{f}_\varphi^n(x) = a\left(\widehat{F}_{\xi^+}^n(x) - \widehat{F}_{\xi^-}^n(x)\right),$$

on en déduit que,

$$\forall x \in \Omega, \widehat{f}_\varphi^n(x) \in a\left(\overline{\mathbb{E}}_{\pi^{x^+}}(E_n) \ominus \overline{\mathbb{E}}_{\pi^{x^-}}(E_n)\right) = \underline{f}_{(\kappa)}^n(x).$$

□

**Propriété III.19.** La seconde inclusion est fautive en général. C'est-à-dire :

$$\forall x \in \Omega, \exists y \in \underline{f}_{(\kappa)}^n(x), \forall \varphi \in \mathcal{D}(a, (\pi^+, \pi^-)), y \neq \widehat{f}_\varphi^n(x). \quad (\text{III.34})$$

Pour s'en convaincre, il suffit, en contre exemple, de considérer le noyau d'Epanechnikov. En effet, lorsqu'on base l'estimation maxitive que nous proposons sur ce noyau, la seule estimation sommative appartenant à l'intervalle obtenu est celle basée sur le noyau d'Epanechnikov défini par l'expression (I.32). Pour montrer cette propriété, nous montrons que la borne inférieure de l'estimation maxitive, en un point  $x \in \Omega$ , est égale à l'estimateur de densité de Parzen-Rosenblatt utilisant le noyau d'Epanechnikov. On en déduit aisément que

$$\forall x \in \Omega, \forall y \in \left] \underline{f}_{(\kappa)}^n(x), \overline{f}_{(\kappa)}^n(x) \right] \nexists \varphi \in \mathcal{D}(a, (\pi^+, \pi^-)) / y = \widehat{f}_{\varphi}^n(x).$$

**Preuve :**  $\mathbf{e}$ , le noyau d'Epanechnikov de largeur de bande égale à 1, est donné (I.32) par :

$$\mathbf{e}(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{\{|x| \leq 1\}}.$$

L'estimateur de Parzen-Rosenblatt  $\widehat{f}_{\mathbf{e}}^n$ , défini par (I.24), est donnée par :

$$\begin{aligned} \widehat{f}_{\mathbf{e}}^n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{e}(x - x_i) \\ &= \frac{3}{4n} \sum_{x \geq x_i} (1 - (x - x_i)^2) + \frac{3}{4n} \sum_{x < x_i} (1 - (x - x_i)^2). \end{aligned}$$

La décomposition sommative de  $\mathbf{e}$ , définie par (III.14), est donnée par :

$$\begin{aligned} -d\mathbf{e}(x) &= a(\eta^+(x) - \eta^-(x)) \\ &= \frac{3}{4} \left( 2x\mathbb{1}_{[0,1]}(x) + 2x\mathbb{1}_{[-1,0]}(x) \right). \end{aligned}$$

Le noyau maxitif  $\pi^+$  (rsp.  $\pi^-$ ) le plus spécifique qui domine  $\eta^+$  (rsp.  $\eta^-$ ) est donnée par :

$$\pi^+(x) = x^2\mathbb{1}_{[0,1]}(x) \text{ et } \pi^-(x) = x^2\mathbb{1}_{[-1,0]}.$$

La borne inférieure  $\underline{F}_{\pi^+}^n$  (rsp. supérieure  $\overline{F}_{\pi^-}^n$ ) de  $\overline{F}_{\pi^+}^n$  (rsp.  $\underline{F}_{\pi^-}^n$ ) est donnée par :

$$\underline{F}_{\pi^+}^n(x) = \frac{1}{n} \left( \sum_{x \geq x_i} 1 - \sum_{x \geq x_i} (x - x_i)^2 \right),$$

et

$$\overline{F}_{\pi^-}^n(x) = \frac{1}{n} \left( \sum_{x > x_i} 1 + \sum_{x < x_i} (x - x_i)^2 \right).$$

La borne inférieure  $\underline{f}_{(\mathbf{e})}^n$  de  $\overline{f}_{(\mathbf{e})}^n$ , définie par (III.32), est donnée par :

$$\begin{aligned} \underline{f}_{(\mathbf{e})}^n(x) &= a \left( \underline{F}_{\pi^+}^n(x) - \overline{F}_{\pi^-}^n(x) \right) \\ &= \frac{3}{4n} \left( 1 - \sum_{x \geq x_i} (x - x_i)^2 - \sum_{x < x_i} (x - x_i)^2 \right) \\ &= \frac{3}{4n} \sum_{x \geq x_i} (1 - (x - x_i)^2) + \frac{3}{4n} \sum_{x < x_i} (1 - (x - x_i)^2) \\ &= \widehat{f}_{\mathbf{e}}^n(x). \end{aligned}$$

Partant de cette propriété, on en déduit aisément que la seule densité appartenant à  $\underline{f}_{(\epsilon)}^n$  est justement  $\hat{f}_\epsilon^n$ . En effet, soit  $g \in \underline{f}_{(\epsilon)}^n$  une fonction de  $L^1(\Omega)$ . Si  $g \neq \hat{f}_\epsilon^n$  alors  $\exists A \subset \Omega$  tel que  $\forall x \in A, g(x) - \hat{f}_\epsilon^n(x) = \epsilon(x) > 0$ . Donc  $\int_\Omega g(x)dx = \int_\Omega \hat{f}_\epsilon^n(x)dx + \int_A \epsilon(x)dx > 1$ . Donc  $g$  n'est pas une densité. □

### III.2.4 Convergence de l'estimation maxitive de la densité

Pour une estimation imprécise, ce qui est généralement souhaité c'est une convergence en valeur et en spécificité. Soyons plus explicite. Nous supposons, pour tout  $x \in \Omega$ , que l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n(x)$  est une enveloppe convexe de toutes les estimations précises  $\hat{f}_{\varphi_\Delta}^n(x)$  avec  $\varphi \in \mathcal{D}(a, (\pi^+, \pi^-))$ . Sachant qu'une des propriétés des estimateurs de Parzen-Rosenblatt est que, si  $\Delta \rightarrow 0$  et  $n\Delta \rightarrow \infty$  quand  $n \rightarrow \infty$ , alors  $\forall x \in \Omega, \hat{f}_{\varphi_\Delta}^n(x) \rightarrow f(x)$ , si cette enveloppe convexe était spécifique, on devrait avoir pour tout  $x \in \Omega$  :

$$\text{si } \Delta \rightarrow 0 \text{ et } n\Delta \rightarrow \infty \text{ quand } n \rightarrow \infty, \text{ alors } \underline{f}_{(\kappa_\Delta)}^n(x) \rightarrow f(x), \quad (\text{III.35})$$

c'est à dire que l'estimation imprécise de la densité tend vers une densité précise. Ce que nous voyons dans la section suivante est que l'imprécision moyenne de  $\underline{f}_{(\kappa_\Delta)}^n$  ne dépend ni de  $\Delta$  ni de  $n$ . Par contre, une certaine convergence reste vraie héritée de la convergence des estimateurs de Parzen-Rosenblatt : la densité vraie appartient à son estimation imprécise lorsque  $\Delta \rightarrow 0$  et  $n \rightarrow \infty$ . Nous appelons cette propriété *la convergence en un certain sens*.

#### III.2.4.1 Imprécision de l'intervalle d'estimation de la densité

Nous avons déjà noté, en section II.3.3.2, que l'imprécision de l'estimation maxitive de la fonction de répartition dépend de la largeur de bande du noyau maxitif considéré mais est indépendante du nombre d'observations. Cette partie montre que l'estimation maxitive de la densité est indépendante de la largeur de bande et du nombre d'observations.

Soit  $\Gamma_{(\kappa_\Delta)}^n$  la fonction définie par :

$$\forall x \in \Omega, \Gamma_{(\kappa_\Delta)}^n(x) = \overline{f}_{(\kappa_\Delta)}^n(x) - \underline{f}_{(\kappa_\Delta)}^n(x), \quad (\text{III.36})$$

où  $\underline{f}_{(\kappa_\Delta)}^n$  (resp.  $\overline{f}_{(\kappa_\Delta)}^n$ ) est la borne inférieure (resp. supérieure) de  $\underline{f}_{(\kappa_\Delta)}^n$ , l'intervalle d'estimation de la densité, défini par (III.31).  $\Gamma_{(\kappa_\Delta)}^n$  définit l'imprécision de  $\underline{f}_{(\kappa_\Delta)}^n$  en chaque point de  $\Omega$ . Par abus de langage, on appellera  $\Gamma_{(\kappa_\Delta)}^n$  la fonction d'imprécision de  $\underline{f}_{(\kappa_\Delta)}^n$ .

Soit  $\rho(\pi_\Delta^+)$  (resp.  $\rho(\pi_\Delta^-)$ ) la granularité de  $\pi_\Delta^+$  (resp.  $\pi_\Delta^-$ ), définie par (II.73). Soit  $\Psi_{\pi_\Delta^+}^n$  (resp.  $\Psi_{\pi_\Delta^-}^n$ ) l'imprécision de  $\overline{F}_{\pi_\Delta^+}^n$  (resp.  $\overline{F}_{\pi_\Delta^-}^n$ ). Le théorème suivant montre que l'intégrale de l'imprécision  $\Gamma_{(\kappa_\Delta)}^n$  de  $\underline{f}_{(\kappa_\Delta)}^n$  est égale à une constante que nous notons  $\zeta(\kappa)$ .

**Théorème III.20 (Intégrale de l'imprécision)** *L'intégrale sur  $\Omega$  de la fonction d'imprécision  $\Gamma_{(\kappa_\Delta)}^n$  de  $\underline{f}_{(\kappa_\Delta)}^n$ , définie par (III.36), est égale à une constante qui ne dépend que de la granularité des noyaux maxitifs utilisés pour dominer les noyaux intervenant dans la décomposition sommative de la dérivée de  $\kappa$ .*



**Preuve :** D'après l'expression (III.32), pour tout  $x \in \Omega$ , on a :

$$\Gamma_{(\kappa_\Delta)}^n(x) = \frac{a}{\Delta} \left( \overline{F}_{\pi_\Delta^+}^n(x) - \underline{F}_{\pi_\Delta^+}^n(x) + \overline{F}_{\pi_\Delta^-}^n(x) - \underline{F}_{\pi_\Delta^-}^n(x) \right). \quad (\text{III.37})$$

En intégrant la fonction  $\Gamma_{(\kappa_\Delta)}^n$  sur  $\Omega$ , on a :

$$\begin{aligned} \zeta(\kappa) &= \int_{\Omega} \Gamma_{(\kappa_\Delta)}^n(x) dx \\ &= \frac{a}{\Delta} \left[ \int_{\Omega} \left( \overline{F}_{\pi_\Delta^+}^n(x) - \underline{F}_{\pi_\Delta^+}^n(x) \right) dx + \int_{\Omega} \left( \overline{F}_{\pi_\Delta^-}^n(x) - \underline{F}_{\pi_\Delta^-}^n(x) \right) dx \right] \\ &= \frac{a}{\Delta} \left( \int_{\Omega} \Psi_{\pi_\Delta^+}^n(x) dx + \int_{\Omega} \Psi_{\pi_\Delta^-}^n(x) dx \right) \text{ par l'expression (II.86)} \\ &= \frac{a}{\Delta} \left( \rho(\pi_\Delta^+) + \rho(\pi_\Delta^-) \right) \text{ par l'expression (II.74)} \\ &= a \left( \rho(\pi^+) + \rho(\pi^-) \right). \end{aligned}$$

Le facteur  $a$  ne dépend que de la fonction de forme  $\kappa$  et la granulosité  $\rho(\pi^+)$  et  $\rho(\pi^-)$  à la fois de  $\kappa$  et de la méthode de domination choisie.  $\zeta(\kappa)$  ne dépend donc ni de  $\Delta$  ni de  $n$ .  $\square$

On a comme corollaire immédiat du théorème III.20 :

**Corollaire III.21.** *Si  $\pi^+$  et  $\pi^-$ , issus de la décomposition et de la domination de  $\kappa$ , sont deux noyaux maxitifs translatés d'un même noyau maxitif  $\pi$ , c'est-à-dire,  $\forall u \in \Omega$ ,  $\pi^+(u) = \pi(u - \frac{1}{2})$  et  $\pi^-(u) = \pi(u + \frac{1}{2})$ , alors*

$$\zeta(\kappa) = 2a\rho(\pi). \quad (\text{III.38})$$

**Preuve :** D'après le corollaire II.73, on a  $\rho(\pi^+) = \rho(\pi^-) = \rho(\pi)$ . D'où l'expression (III.38).  $\square$

Le résultat du théorème III.20, nous permet d'affirmer que l'intervalle d'estimation de la densité  $\overline{f}_{(\kappa_\Delta)}^n$  ne converge pas vers la vraie densité. C'est-à-dire,  $\forall x \in \Omega$  :

$$\text{si } \Delta \rightarrow 0 \text{ et } n\Delta \rightarrow \infty \text{ quand } n \rightarrow \infty, \text{ alors } \Gamma_{(\kappa_\Delta)}^n(x) = \left( \overline{f}_{(\kappa_\Delta)}^n(x) - \underline{f}_{(\kappa_\Delta)}^n(x) \right) \not\rightarrow 0. \quad (\text{III.39})$$

### III.2.4.2 Convergence en un certain sens

L'estimation maxitive ne peut converger au sens de l'expression (III.35) puisque son imprécision ne dépend ni de la largeur de bande du noyau choisi, ni du nombre d'observations utilisé. Elle vérifie cependant la propriété suivante :

$$\text{si } \Delta \rightarrow 0 \text{ quand } n \rightarrow \infty, \text{ alors } \forall x \in \Omega, f(x) \in \overline{f}_{(\kappa_\Delta)}^n(x), \quad (\text{III.40})$$

que nous appelons convergence *en un certain sens*.

Pour prouver cette propriété, définissons tout d'abord la *densité convoluée*, notée  $f_\kappa$ , comme la convolution de la vraie densité  $f$  avec le noyau sommatif  $\kappa$  par :

$$\forall x \in \Omega, f_\kappa(x) = (f \star \kappa)(x) = \int_{\Omega} f(u)\kappa(x-u)du. \quad (\text{III.41})$$

L'estimation maxitive de la densité  $\overline{f}_{(\kappa)}^n$ , définie par (III.31), peut être vue comme un ensemble de fonctions défini comme suit :

$$\overline{f}_{(\kappa)}^n = \left\{ f \in L^1(\Omega) : \forall x \in \Omega, f(x) \in \overline{f}_{(\kappa)}^n(x) = [\underline{f}_{(\kappa)}^n(x), \overline{f}_{(\kappa)}^n(x)] \right\}.$$

$\overline{f}_{(\kappa)}$  la limite simple de  $\overline{f}_{(\kappa)}^n$  formée par la limite simple de ses bornes inférieures et supérieures lorsque  $n$  tend vers  $\infty$ ,  $\overline{f}_{(\kappa)}$  vérifie la propriété suivante :

**Propriété III.22 (Inclusion de la densité convoluée)**  $\forall \kappa \in \mathcal{K}_1(\Omega)$ , on a :

$$\forall x \in \Omega, f_{\kappa}(x) \in \overline{f}_{(\kappa)}(x). \quad (\text{III.42})$$

**Preuve :** Cette preuve s'appuie sur un résultat issu de ([104], pages 36-37) : l'estimateur de densité  $\widehat{f}_{\kappa}^n$  converge simplement vers la densité convoluée  $f_{\kappa}$  :

$$\forall x \in \Omega, \widehat{f}_{\kappa}^n(x) \xrightarrow{n \rightarrow \infty} f_{\kappa}(x).$$

Or, d'après le théorème III.18, on a :

$$\forall n \in \mathbb{N}, \widehat{f}_{\kappa}^n(x) \in \overline{f}_{(\kappa)}^n(x).$$

Si  $n$  tends vers  $\infty$ , cette inclusion reste vraie. □

Lorsque nous avons abordé les propriétés des noyaux sommatifs, nous avons vu que, si  $\kappa$  était un noyau sommatif, alors  $\kappa_{\Delta} \xrightarrow{\Delta \rightarrow 0} \delta$ , où  $\delta$  est l'hypothétique fonction associée à la distribution de Dirac (voir aussi [72] Théorème 5). La distribution de Dirac pouvant être vue comme une limite, au sens des distributions, d'un noyau sommatif lorsque sa largeur de bande tend vers 0, on peut noter  $\overline{f}_{(\delta)}$  la limite de  $\overline{f}_{(\kappa_{\Delta})}$ , quand  $\Delta$  tend vers 0. On a alors la propriété suivante :

**Propriété III.23 (Inclusion de la vraie densité)**  $\forall \kappa \in \mathcal{K}_1(\Omega)$ , on a :

$$\forall x \in \Omega, f(x) \in \lim_{\Delta \rightarrow 0} \overline{f}_{(\kappa_{\Delta})}(x) = \overline{f}_{(\delta)}(x). \quad (\text{III.43})$$

**Preuve :** En nous basant sur [104] :

$$\text{si } \Delta \rightarrow 0 \text{ quand } n \rightarrow \infty, \text{ alors } \forall x \in \Omega, f_{\kappa_{\Delta}}(x) \rightarrow f(x),$$

or, d'après la propriété III.22, on a :

$$\forall x \in \Omega, f_{\kappa_{\Delta}}(x) \in \overline{f}_{(\kappa_{\Delta})}(x).$$

Si  $\Delta \rightarrow 0$  quand  $n \rightarrow \infty$ , cette inclusion reste vraie. □

### III.2.4.3 Comment expérimenter la convergence d'une estimation imprécise de la densité ?

La convergence de l'estimateur de densité vers la vraie densité est prouvée grâce à la distance induite par la norme de l'espace  $L^1(\Omega)$  [27, 104]. Nous avons vu, en section précédente, que le comportement asymptotique, quand  $n$  tends vers l'infini, montre l'inclusion de la densité convoluée  $f_\kappa$ , définie par (III.41), dans l'intervalle d'estimation  $\underline{f}_{(\kappa)}^n$  défini par (III.31). Nous voyons dans cette section plusieurs méthodes pour approximer la distance entre l'estimation maxitive  $\underline{f}_{(\kappa)}^n$ , qui est une fonction imprécise, et la densité convoluée  $f_\kappa$ , qui est une fonction précise. En effet, cette distance est difficile à calculer dans la pratique. Nous proposons de contourner en partie cette difficulté en nous appuyant sur la définition des bornes inférieures et supérieures de cette distance.

Rappelons la définition de  $d_1(f, g)$ , la distance  $L^1(\Omega)$  entre deux fonctions  $f$  et  $g$  de  $L^1(\Omega)$  :

$$d_1(f, g) = \|f - g\|_{L^1(\Omega)} = \int_{\Omega} |f(x) - g(x)| dx. \quad (\text{III.44})$$

Il est possible d'étendre cette distance pour définir une distance entre l'intervalle d'estimation de la densité  $\underline{f}_{(\kappa)}^n$  et la densité convoluée  $f_\kappa$  par :

$$d(f_\kappa, \underline{f}_{(\kappa)}^n) = \inf_{f \in \mathbb{F}^n} d_1(f_\kappa, f), \quad (\text{III.45})$$

où  $\mathbb{F}^n = \left\{ f \in \underline{f}_{(\kappa)}^n : \int_{\Omega} f(x) dx = 1 \right\}$  est l'ensemble des densités de  $\underline{f}_{(\kappa)}^n$ .  $\underline{f}_{(\kappa)}^n$  est un cas particulier d'ensemble convexe de fonction. Dans ce qui suit nous considérons des ensembles convexes de fonctions. Soit  $\underline{g}$  un tel ensemble, une fonction  $h$  définie par ses deux fonctions limites  $\underline{g}$  et  $\overline{g}$ , appartient à  $\underline{g}$  si  $\forall x \in \Omega$ ,  $h(x) \in \underline{g}(x) = [\underline{g}(x), \overline{g}(x)]$ . Ce qui s'écrit :

$$\underline{g} = \left\{ g \in L^1(\Omega) : \forall x \in \Omega, g(x) \in \underline{g}(x) = [\underline{g}(x), \overline{g}(x)] \right\}. \quad (\text{III.46})$$

Cette distance est particulièrement difficile à calculer car  $\mathbb{F}^n$  n'est pas très bien défini. Nous proposons d'encadrer sa valeur. Pour encadrer  $d(f_\kappa, \underline{f}_{(\kappa)}^n)$ , nous proposons d'utiliser trois extensions de la distance  $d_1$ . Le premier de ces extensions, appelée la *distance de Hausdorff* [82], est définie par :

$$d_H(h, \underline{g}) = \int_{\Omega} \sup_{g \in \underline{g}} |h(x) - g(x)| dx, \quad (\text{III.47})$$

où  $h$  est une fonction de  $L^1(\Omega)$ .

Puisque  $\underline{g}$  est un intervalle de fonctions, la *distance de Hausdorff*, définie par (III.47), peut se réécrire comme suit :

$$d_H(h, \underline{g}) = \int_{\Omega} \max \left\{ \left| \underline{g}(x) - h(x) \right|, \left| \overline{g}(x) - h(x) \right| \right\} dx. \quad (\text{III.48})$$

La distance de Hausdorff est souvent utilisée pour comparer deux sous-ensembles d'un espace métrique sous-jacent [77]. On utilise ici cette distance pour comparer un objet imprécis  $\underline{g}$ , avec un objet précis  $h$ . La valeur minimale qui peut atteindre cette distance pour  $\underline{g}$  donné :

$$\beta = \frac{1}{2} \int_{\Omega} (\overline{g}(x) - \underline{g}(x)) dx. \quad (\text{III.49})$$

Cette distance atteint sa valeur en  $h = \tilde{g}$ ,  $\tilde{g}$  étant la fonction médiane de  $\underline{g}$  définie par :

$$\forall x \in \Omega, \tilde{g}(x) = \frac{1}{2} (\overline{g}(x) + \underline{g}(x)). \quad (\text{III.50})$$

On définit la *distance médiane*, via la distance  $L^1(\Omega)$ , par :

$$d_M(h, \underline{g}) = d_1(h, \tilde{g}). \quad (\text{III.51})$$

Enfin, nous définissons l'*extension naturelle* de  $d_1$  par :

$$d_S(h, \underline{g}) = \int_{\Omega} \inf_{g \in \underline{g}} |h(x) - g(x)| dx. \quad (\text{III.52})$$

Puisque  $\underline{g}$  est un intervalle des fonctions, l'*extension naturelle* de la distance  $L^1(\Omega)$ , définie par (III.52), peut se réécrire comme suit :

$$d_S(h, \underline{g}) = \int_{\Omega} \left( |h(x) - \underline{g}(x)| \mathbb{1}_{\{h(x) \leq \underline{g}(x)\}} + |\overline{g}(x) - h(x)| \mathbb{1}_{\{h(x) \geq \overline{g}(x)\}} \right) dx. \quad (\text{III.53})$$

Notons que, si  $\forall x \in \Omega, h(x) \in \underline{g}(x)$ , alors  $d_S(h, \underline{g}) = 0$ .

**Propriété III.24 (Relation entre les distances)** La distance entre l'intervalle d'estimation  $\underline{f}_{(\kappa)}^n$  et la densité convoluée  $f_{\kappa}$  est comprise entre l'extension naturelle de la distance  $L^1(\Omega)$ ,  $d_S$  définie par (III.53) et la distance de Hausdorff  $d_H$  définie par (III.48) :

$$d_S(f_{\kappa}, \underline{f}_{(\kappa)}^n) \leq d(f_{\kappa}, \underline{f}_{(\kappa)}^n) \leq d_H(f_{\kappa}, \underline{f}_{(\kappa)}^n). \quad (\text{III.54})$$

**Preuve :** Par construction, on a  $\mathbb{F}^n \subseteq \underline{f}_{(\kappa)}^n$ . Donc,

$$\inf_{g \in \underline{f}_{(\kappa)}^n} d_1(f_{\kappa}, g) \leq \inf_{g \in \mathbb{F}^n} d_1(f_{\kappa}, g) \leq \sup_{g \in \mathbb{F}^n} d_1(f_{\kappa}, g) \leq \sup_{g \in \underline{f}_{(\kappa)}^n} d_1(f_{\kappa}, g),$$

ce qui donne directement,

$$\inf_{g \in \underline{f}_{(\kappa)}^n} d_1(f_{\kappa}, g) \leq d(f_{\kappa}, \underline{f}_{(\kappa)}^n) \leq \sup_{g \in \underline{f}_{(\kappa)}^n} d_1(f_{\kappa}, g).$$

Remarquons que,

$$\int_{\Omega} \inf_{g \in \underline{f}_{(\kappa)}^n} |f_{\kappa}(x) - g(x)| dx \leq \inf_{g \in \underline{f}_{(\kappa)}^n} \int_{\Omega} |f_{\kappa}(x) - g(x)| dx,$$

et

$$\sup_{g \in \underline{f}_{(\kappa)}^n} \int_{\Omega} |f_{\kappa}(x) - g(x)| dx \leq \int_{\Omega} \sup_{g \in \underline{f}_{(\kappa)}^n} |f_{\kappa}(x) - g(x)| dx.$$

On en déduit,

$$d_S(f_{\kappa}, \underline{f}_{(\kappa)}^n) \leq \inf_{g \in \underline{f}_{(\kappa)}^n} d_1(f_{\kappa}, g),$$

et

$$\sup_{g \in \underline{f}_{(\kappa)}^n} d_1(f_{\kappa}, g) \leq d_H(f_{\kappa}, \underline{f}_{(\kappa)}^n).$$

Donc, finalement,

$$d_S(f_\kappa, \underline{f}_{(\kappa)}^n) \leq d(f_\kappa, \overline{f}_{(\kappa)}^n) \leq d_H(f_\kappa, \underline{f}_{(\kappa)}^n).$$

□

Dans le but d'expérimenter cette convergence, nous introduisons la *distance médiane* notée  $d_M$ . Cette distance est définie, via la distance  $L^1(\Omega)$ , par :

$$d_M(h, \underline{g}) = d_1(h, \tilde{g}), \quad (\text{III.55})$$

où  $\tilde{g}$  est la fonction médiane de  $\underline{g}$  définie par :

$$\forall x \in \Omega, \tilde{g}(x) = \frac{1}{2}(\overline{g}(x) + \underline{g}(x)). \quad (\text{III.56})$$

On peut donc définir la distance médiane définie entre l'intervalle d'estimation  $\overline{f}_{(\kappa_\Delta)}^n$  et la densité convoluée  $f_{\kappa_\Delta}$  par :

$$d_M(f_{\kappa_\Delta}, \overline{f}_{(\kappa_\Delta)}^n) = d_1(f_{\kappa_\Delta}, \tilde{f}_{\kappa_\Delta}), \quad (\text{III.57})$$

où  $\tilde{f}_{\kappa_\Delta}$  est la fonction médiane de  $\overline{f}_{(\kappa_\Delta)}^n$  définie par :

$$\forall x \in \Omega, \tilde{f}_{\kappa_\Delta}(x) = \frac{1}{2}(\overline{f}_{\kappa_\Delta}(x) + \underline{f}_{\kappa_\Delta}(x)). \quad (\text{III.58})$$

### III.3 Spécificité de l'estimation maxitive de la densité

Une des causes principales de la perte de spécificité est l'utilisation de l'extension aux intervalles de la soustraction  $\ominus$ . Dans cette partie, nous proposons d'éliminer cette cause en considérant des noyaux particuliers. Dans un premier temps nous considérons les noyaux sommatifs monomodaux symétriques dont la dérivée peut s'exprimer comme la translation d'un seul noyau développé au paragraphe III.1.4.2. Dans ce cas, la complexité de l'algorithme de calcul de l'extension maxitive de densité est équivalente à celle du cas général. Dans un second temps nous supposons simplement que la disparition de l'opérateur de Minkowski se fait au détriment de la complexité du calcul de l'estimation maxitive de densité qui augmente sensiblement.

#### III.3.1 Première restriction sur l'intervalle d'estimation

Dans cette première restriction, nous considérons que le noyau sommatif de base  $\kappa$  est tel que la décomposition canonique de sa dérivée ne fait intervenir qu'un seul noyau sommatif  $\eta$  sous la forme :

$$\forall u \in \Omega, -d\kappa(u) = a\left(\eta\left(u - \frac{1}{2}\right) - \eta\left(u + \frac{1}{2}\right)\right),$$

avec  $a = \int_\Omega \max(0, -d\kappa(u))du$  et  $\eta \in \mathcal{K}(\Omega)$ .

La technique de domination du noyau sommatif permet de définir le noyau maxitif  $\pi$  le plus spécifique qui domine  $\eta$ . On peut donc définir un ensemble des noyaux sommatifs que nous notons  $\mathcal{D}(a, \pi)$  par :

$$\mathcal{D}(a, \pi) = \left\{ \begin{array}{l} v \in \mathcal{K}_1(\Omega), \exists \xi \in \mathcal{M}(\pi), \text{ tel que} \\ \forall u \in \Omega, -dv(u) = a \left( \xi(u - \frac{1}{2}) - \xi(u + \frac{1}{2}) \right) \end{array} \right\}, \quad (\text{III.59})$$

où  $\mathcal{M}(\pi)$  est la famille des noyaux sommatifs dominés par le noyau maxitif  $\pi$ .

Cet ensemble est naturellement non vide car, par construction,  $\kappa \in \mathcal{D}(a, \pi)$ . On peut alors construire l'intervalle d'estimation de la densité, noté  $\underline{f}_{[\kappa]}^n$ , en utilisant le noyau maxitif translaté  $\pi$  par :

$$\underline{f}_{[\kappa]}^n(x) = a \left( \underline{\mathbb{E}}_{\pi^x}(E_n^+ - E_n^-) \right), \quad (\text{III.60})$$

où  $E_n^+$  et  $E_n^-$  sont deux fonctions de répartition empirique définies dans (III.22) et  $\underline{\mathbb{E}}_{\pi}(\cdot)$  est l'espérance imprécise basée sur le noyau maxitif  $\pi$ .

### Algorithme de calcul de bornes

Nous avons vu, en section III.2.2, que l'algorithme de calcul de l'intervalle d'estimation de la densité  $\underline{f}_{(\kappa)}^n$ , défini par (III.31), est directe. Dans cette partie, nous présentons l'algorithme de calcul de  $\underline{f}_{[\kappa]}^n$  défini par (III.60). La nature de l'intégrale de Choquet utilisée pour calculer  $\underline{f}_{[\kappa]}^n$  nécessite une adaptation. Cette adaptation passe par une reformulation des éléments intervenant dans ce calcul d'espérance imprécise basée sur le noyau maxitif  $\pi$ .

En posant  $\Delta E_n = E_n^+ - E_n^-$ , l'intervalle d'estimation  $\underline{f}_{[\kappa]}^n$ , défini par (III.60), peut se réécrire comme suit :

$$\forall x \in \Omega, \underline{f}_{[\kappa]}^n(x) = a \underline{\mathbb{E}}_{\pi^x}(\Delta E_n) = a \left[ \mathbb{C}_{N_{\pi^x}}(\Delta E_n), \mathbb{C}_{\Pi_{\pi^x}}(\Delta E_n) \right], \quad (\text{III.61})$$

où  $\mathbb{C}_{\Pi_{\pi}}(\Delta E_n)$  (rsp.  $\mathbb{C}_{N_{\pi}}(\Delta E_n)$ ) est l'intégrale de Choquet de  $\Delta E_n$  par rapport à la mesure de possibilité  $\Pi_{\pi}$  (rsp. la mesure de nécessité  $N_{\pi}$ ) définie par (II.3).

La fonction  $\Delta E_n$  est une *fonction en escalier*<sup>4</sup> (*step function* en anglais). Nous pouvons l'écrire :

$$\begin{aligned} \Delta E_n(u) &= \frac{1}{n} \sum_{i=1}^n \left( H(u - x_i + \frac{1}{2}) - H(u - x_i - \frac{1}{2}) \right), \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[-\frac{1}{2}, +\frac{1}{2}[}(u - x_i), \end{aligned} \quad (\text{III.62})$$

où  $\mathbb{1}_A$  est la fonction indicatrice de  $A \subseteq \Omega$ , définie par (I.3) et  $H$  est la distribution de Heaviside définie par (I.40).

Par construction, la fonction en escalier  $\Delta E_n$  a  $(2n+1)$  valeurs différentes. Considérons  $\{w_i\}_{i \in \{0, \dots, 2n+1\}}$  l'ensemble des  $(2n+2)$  valeurs provenant de l'ensemble des observations  $\{x_i\}_{i \in \{1, \dots, n\}}$  défini par :

$$w_i = \begin{cases} x_i - \frac{1}{2}, & \text{si } i \in \{1, \dots, n\}, \\ x_{i-n} + \frac{1}{2}, & \text{si } i \in \{n+1, \dots, 2n\}, \end{cases} \quad (\text{III.63})$$

---

4. Puisqu'il s'agit d'une différence de deux fonctions en escalier.

où  $w_0 = e_{min}$  et  $w_{2n+1} = e_{max}$ .

Par construction, la fonction  $\Delta E_n$  change de valeur en chaque  $\{w_i\}_{i \in \{0, \dots, 2n+1\}}$ . Soit  $\Theta = \{0, \dots, 2n\}$  et soit  $(\cdot)$  la permutation des indices sur l'ordre des  $w_i$  de sorte que  $w_{(1)} \leq w_{(2)} \leq \dots \leq w_{(2n+1)}$ . Soit  $\{W_i\}_{i \in \Theta}$  les  $(2n+1)$  intervalles définis à partir de  $w_i$  par :  $W_i = \{[w_{(i)}, w_{(i+1)}]\}_{i \in \Theta}$ .  $\Delta E_n$  est constante sur chaque intervalle  $W_i$ . Soit  $D_n^i$  la valeur constante de la fonction  $\Delta E_n$  quand  $x \in W_i$  et soit  $z_i$  la valeur médiane de  $W_i$  définie par :  $z_i = \frac{w_{(i)} - w_{(i+1)}}{2}$ , on a :

$$D_n^i = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{[x_k - \frac{1}{2}, x_k + \frac{1}{2}[}(z_i), \quad (\text{III.64})$$

avec  $D_n^0 = D_n^{2n} = 0$ , par construction.

La fonction  $\Delta E_n$  peut donc être réécrite comme suit :

$$\forall u \in \Omega, \Delta E_n(u) = \sum_{i=0}^{2n} D_n^i \mathbb{1}_{W_i}(u). \quad (\text{III.65})$$

D'après l'expression (III.61), le calcul de  $\bar{f}_{[k]}^n$  fait intervenir deux intégrales de Choquet  $\mathbb{C}_{\Pi_\pi}(\Delta E_n)$  et  $\mathbb{C}_{N_\pi}(\Delta E_n)$ . Aux vues de la nature de la fonction  $\Delta E_n$  à intégrer, l'intégrale de Choquet continue peut être remplacée par une intégrale de Choquet discrète par rapport à une distribution de possibilité associée aux  $\{W_i\}_{i \in \Theta}$ . La définition de cette distribution de possibilité discrète sur  $\Theta$  fait l'objet du lemme suivant :

**Lemme III.25 (Distribution de possibilité discrète)** *Une distribution de possibilité continue  $\pi$  sur  $\Omega$ , induit une distribution de possibilité discrète  $\mu$  sur  $\Theta$ , définie par :*

$$\forall k \in \Theta, \mu_k = \Pi_\pi(W_k) = \sup_{x \in W_k} \{\pi(x)\}. \quad (\text{III.66})$$

**Preuve :** Il s'agit de montrer que  $\mu_k$  définie par (III.66) est bien une distribution de possibilité sur  $\Theta$ . Soit  $\mathbf{m}$ , le mode de  $\pi$ , c'est-à-dire la valeur telle que  $\pi(\mathbf{m}) = 1$ . Comme  $\{W_k\}_{k \in \Theta}$  est une partition de  $\Omega$ , il existe un et un seul  $k$  tel que  $\mu_k = 1$ . Donc  $\mu_k$  est une distribution de possibilité monomodale sur  $\Theta$ . □

D'après le lemme III.25, pour tout ensemble  $A \subseteq \Theta$ , on a :

$$\Pi_\mu(A) = \sup_{k \in A} \{\mu^k\} = \sup_{k \in A} \left\{ \Pi_\pi(W_k) \right\} = \Pi_\pi \left( \bigcup_{k \in A} W_k \right).$$

Le passage de la forme continue à la forme discrète pour les deux intégrales de Choquet  $\mathbb{C}_{\Pi_\pi}(\Delta E_n)$  et  $\mathbb{C}_{N_\pi}(\Delta E_n)$  est donnée par le théorème suivant :

**Théorème III.26 (Expression discrète de l'intégrale de Choquet continue)** *Soit  $\Delta E_n$  la fonction d'étape définie par (III.62) et  $D_n$  sa valeur définie par (III.64). Soit  $(\cdot)$  la permutation des indices sur  $\{D_n^i\}_{i \in \Theta}$  de sorte que  $D_n^{(0)} \leq D_n^{(1)} \leq \dots \leq D_n^{(2n)}$  et soit  $\{A_i\}_{i \in \Theta}$  les  $(2n+1)$  sous-ensembles de  $\Theta$  définis par :  $A_{(i)} = \{(i), \dots, (2n)\}$ . L'intégrale de Choquet  $\mathbb{C}_{\Pi_\pi}$  (resp.  $\mathbb{C}_{N_\pi}$ ) de  $\Delta E_n$  par rapport à la mesure de possibilité  $\Pi_\pi$  (resp. nécessité  $N_\pi$ ), est donnée par :*

$$\mathbb{C}_{\Pi_\pi}(\Delta E_n) = \mathbb{C}_{\Pi_\mu}(D_n) = \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) \Pi_\mu(A_{(i)}), \quad (\text{III.67})$$

et

$$\mathbb{C}_{N_\pi}(\Delta E_n) = \mathbb{C}_{N_\mu}(D_n) = \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) N_\mu(A_{(i)}), \quad (\text{III.68})$$

où  $\pi$  est une distribution de possibilité continue sur  $\Omega$  et  $\mu$  est une distribution de possibilité discrète sur  $\Theta$ , définie à partir de  $\pi$  par (III.66).

**Preuve :** D'après l'expression (II.19), on a :

$$\mathbb{C}_{\Pi_\pi}(\Delta E_n) = (C) \int_{\Omega} \Delta E_n d\Pi_\pi = \int_0^\infty \Pi_\pi \left( \left\{ u \in \Omega / \Delta E_n(u) \geq \ell \right\} \right) d\ell. \quad (\text{III.69})$$

Comme  $\Delta E_n$  est une fonction en escalier ayant au plus  $(2n + 1)$  valeurs ordonnées par la permutation  $(\cdot)$ , l'expression (III.69) peut être décomposée en somme de  $(2n + 1)$  intégrales :

$$\begin{aligned} \mathbb{C}_{\Pi_\pi}(\Delta E_n) &= \int_0^{D_n^{(0)}} \Pi_\pi \left( \left\{ u \in \Omega / \Delta E_n(u) \geq D_n^{(0)} \right\} \right) d\ell \\ &+ \int_{D_n^{(0)}}^{D_n^{(1)}} \Pi_\pi \left( \left\{ u \in \Omega / \Delta E_n(u) \geq D_n^{(1)} \right\} \right) d\ell \\ &+ \cdots + \int_{D_n^{(2n-1)}}^{D_n^{(2n)}} \Pi_\pi \left( \left\{ u \in \Omega / \Delta E_n(u) \geq D_n^{(2n)} \right\} \right) d\ell. \end{aligned}$$

Cette somme peut être réduite en  $(2n)$  termes, car  $D_n^{(0)} = 0$ .

En remarquant que

$$\left\{ u \in \Omega / \Delta E_n(u) \geq D_n^{(i)} \right\} = \bigcup_{k=i}^{2n} W_{(k)},$$

il vient

$$\begin{aligned} \mathbb{C}_{\Pi_\pi}(\Delta E_n) &= \int_0^{D_n^{(1)}} \Pi_\pi \left( W_{(1)} \cup \cdots \cup W_{(2n)} \right) d\ell + \cdots + \int_{D_n^{(2n-1)}}^{D_n^{(2n)}} \Pi_\pi \left( W_{(2n)} \right) d\ell \\ &= \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) \Pi_\pi \left( \bigcup_{k=i}^{2n} W_{(k)} \right) \\ &= \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) \sup_{k \in \{i, \dots, 2n\}} \Pi_\pi \left( W_{(k)} \right) \\ &= \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) \Pi_\mu \left( A_{(i)} \right). \end{aligned}$$

De la même manière, on obtient :

$$\mathbb{C}_{N_\pi}(\Delta E_n) = \mathbb{C}_{N_\mu}(D_n) = \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) N_\mu \left( A_{(i)} \right).$$

□

Le lemme suivant découle du théorème III.26.



**Lemme III.27 (Calcul de l'intégrale de Choquet vis à vis de la mesure de nécessité)**

L'intégrale de Choquet définie par l'expression (III.68), calculée par rapport à la mesure de nécessité  $N_\mu$ , peut être calculée par la même procédure de l'expression (III.67), en considérant la permutation  $\sigma(i)$  sur  $\{D_n^i\}_{i \in \Theta}$  de sorte que  $D_n^{\sigma(0)} \geq D_n^{\sigma(1)} \geq \dots \geq D_n^{\sigma(2n)}$ .

**Preuve :** Soit  $B_{(i)} = \{(0), \dots, (i-1)\}$ . Soient  $\sigma(i)$  la permutation sur  $\{D_n^i\}_{i \in \Theta}$  de sorte que  $D_n^{(0)} \geq D_n^{(1)} \geq \dots \geq D_n^{(2n)}$  et  $(\cdot)$  la permutation des indices sur  $\Omega$  définie précédemment. Soit  $A_{\sigma(i)} = \{\sigma(i), \dots, \sigma(2n)\}$ , où  $\sigma(i) = (2n - i + 1)$ . Par construction, on a  $N_\mu(A_{(i)}) = 1 - \Pi_\mu(B_{(i)})$ , donc l'expression (III.68) peut se réécrire comme suit :

$$\begin{aligned} \mathbb{C}_{N_\pi}(\Delta E_n) &= \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) \left( 1 - \Pi_\mu(B_{(i)}) \right) \\ &= \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) - \sum_{i=1}^{2n} \left( D_n^{(i)} - D_n^{(i-1)} \right) \left( \Pi_\mu(B_{(i)}) \right) \\ &= \sum_{i=1}^{2n} D_n^{(i-1)} \left( \Pi_\mu(B_{(i)}) - \Pi_\mu(B_{(i-1)}) \right) \text{ car } D_n^{(0)} = 0 \text{ et } \Pi(B_{(2n)}) = 1 \\ &= \sum_{i=1}^{2n} \left( D_n^{\sigma(i)} - D_n^{\sigma(i-1)} \right) \Pi_\mu(A_{\sigma(i)}). \end{aligned}$$

Donc :

$$\mathbb{C}_{N_\pi}(\Delta E_n) = \mathbb{C}_{N_\mu}(D_n) = \sum_{i=1}^{2n} \left( D_n^{\sigma(i)} - D_n^{\sigma(i-1)} \right) \Pi_\mu(A_{\sigma(i)}). \quad (\text{III.70})$$

□

Le résultat suivant découle immédiatement du lemme III.25.

**Lemme III.28 (Distribution de possibilité discrète translatée)** La distribution de possibilité continue  $\pi^x$  obtenue par la translation en  $x \in \Omega$  de la distribution de possibilité  $\pi$  sur  $\Omega$ , induit une distribution de possibilité discrète et translatée  $\mu^x$  sur  $\Theta$  définie par :

$$\forall k \in \Theta, \mu_k^x = \Pi_{\pi^x}(W_k) = \sup_{u \in W_k} \{\pi^x(u)\}. \quad (\text{III.71})$$

D'après le lemme III.28 et le théorème III.26, l'espérance imprécise  $\overline{\mathbb{E}}_{\pi^x}(\Delta E_n)$  de la fonction  $\Delta E_n$  basée sur le noyau maxitif translaté  $\pi$ , définie par (III.61), peut être alors réécrite comme suit :

$$\overline{\mathbb{E}}_{\pi^x}(\Delta E_n) = [\mathbb{C}_{N_{\pi^x}}(\Delta E_n), \mathbb{C}_{\Pi_{\pi^x}}(\Delta E_n)] = [\mathbb{C}_{N_{\mu^x}}(D_n), \mathbb{C}_{\Pi_{\mu^x}}(D_n)]. \quad (\text{III.72})$$

Les calculs de  $\{\overline{f}_{[\kappa]}^n(y_j)\}_{j \in \{1, \dots, p\}}$ , sont évalués sur les points  $\{y_j\}_{j \in \{1, \dots, p\}}$  (cf. section III.2.2). Ces calculs peuvent être résumés en deux étapes :

**Etape 1 :** Calculs de  $\{W_i\}_{i \in \{0, \dots, 2n\}}$  et  $\{D_n^i\}_{i \in \{0, \dots, 2n\}}$  qui nécessitent les observations  $\{x_i\}_{i \in \{1, \dots, n\}}$ .

**Etape 2 :** Calcul, pour chaque  $\{y_j\}_{j \in \{1, \dots, p\}}$ , de l'intervalle d'estimation de la densité  $\{\overline{f}_{[\kappa]}^n(y_j)\}_{j \in \{1, \dots, p\}}$ . Il nécessite le noyau maxitif  $\pi$ , les intervalles  $\{W_i\}_{i \in \{0, \dots, 2n\}}$  et les constantes  $\{D_n^i\}_{i \in \{0, \dots, 2n\}}$ .

**Data:** les observations  $\{x_i\}_{i \in \{1, \dots, n\}}$ .  
**Result:** les intervalles  $\{W_i\}_{i \in \{0, \dots, 2n\}}$ , les valeurs discrètes  $\{D_n^i\}_{i \in \{0, \dots, 2n\}}$ .  
**begin**  
  Calculer l'ensemble de  $(2n + 1)$  valeurs  $\{w_i\}_{i \in \{0, \dots, 2n+1\}}$  en utilisant l'expression (III.63) avec  $w_0 = e_{min}$ ,  $w_{2n+1} = e_{max}$ .  
  Trier les  $\{w_i\}_{i \in \{0, \dots, 2n+1\}}$  par ordre croissant.  
  Calculer les  $(2n + 1)$  intervalles  $\{W_i\}_{i \in \{0, \dots, 2n\}}$ ,  
 $W_i = [w_{(i)}, w_{(i+1)}[$  si  $i \in [0, 2n - 1]$ ,  $W_{2n} = [w_{(2n)}, w_{(2n+1)}]$ .  
  Calculer les  $(2n + 1)$  centres de  $W_i$  par :  $z_i = \frac{w_{(i)} + w_{(i+1)}}{2}$ .  
  Calculer les valeurs constantes  $\{D_n^i\}_{i \in \{0, \dots, 2n\}}$  par l'expression (III.64).  
**end**

**Algorithm 3:** Calcul de  $\{W_i\}_{i \in \{0, \dots, 2n\}}$  et  $\{D_n^i\}_{i \in \{0, \dots, 2n\}}$ .

Les deux algorithmes suivants décrivent plus formellement le processus de calcul.

**Data:** les points  $\{y_j\}_{j \in \{1, \dots, p\}}$ , le noyau maxitif  $\pi$ , les intervalles  $\{W_i\}_{i \in \{0, \dots, 2n\}}$ , les valeurs  $\{D_n^i\}_{i \in \{0, \dots, 2n\}}$  et le facteur  $a$ .  
**Result:**  $\left\{ \left[ \underline{f}_{[\kappa]}^n(y_j), \overline{f}_{[\kappa]}^n(y_j) \right] \right\}_{j \in \{1, \dots, p\}}$ .  
**begin**  
  Trier les  $\{D_n^i\}$  par ordre croissant.  
  Appliquer la même permutation sur les  $\{W_i\}$ .  
  **for**  $j = 1$  **to**  $p$  **do**  
    Calculer les  $\{\mu^i\}_{i \in \{0, \dots, 2n\}}$  par :  $\mu^i = \Pi_{\pi^{y_j}}(W_i)$  (expression (III.66))  
     $\Pi \leftarrow 0$ ,  $\overline{f}_{[\kappa]}^n(y_j) \leftarrow 0$   
    **for**  $i = 2n$  **to**  $1$  **down to**  $1$  **do**  
       $\Pi = \max(\Pi, \mu^i)$   
       $\overline{f}_{[\kappa]}^n(y_j) = \overline{f}_{[\kappa]}^n(y_j) + (D_n^i - D_n^{i-1})\Pi$   
    **end**  
     $\overline{f}_{[\kappa]}^n(y_j) = a \overline{f}_{[\kappa]}^n(y_j)$   
     $\Pi \leftarrow 0$ ,  $\underline{f}_{[\kappa]}^n(y_j) \leftarrow 0$   
    **for**  $i = 1$  **to**  $2n$  **do**  
       $\Pi = \max(\Pi, \mu^i)$   
       $\underline{f}_{[\kappa]}^n(y_j) = \underline{f}_{[\kappa]}^n(y_j) + (D_n^i - D_n^{i-1})\Pi$   
    **end**  
     $\underline{f}_{[\kappa]}^n(y_j) = a \underline{f}_{[\kappa]}^n(y_j)$   
  **end**  
**end**

**Algorithm 4:** Calcul de  $\left\{ \left[ \underline{f}_{[\kappa]}^n(y_j), \overline{f}_{[\kappa]}^n(y_j) \right] \right\}_{j \in \{1, \dots, p\}}$ .

### Relation avec l'estimateur de densité de Parzen-Rosenblatt

Comme nous l'avons vu en section III.2.3, l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_{\kappa}^n$ , défini par (ref), est inclus dans l'intervalle d'estimation de la densité  $\overline{f}_{[\kappa]}^n$ , défini par

(III.31). Nous montrons dans cette partie que l'estimateur  $\hat{f}_\kappa^n$  est inclus dans l'intervalle d'estimation  $\underline{f}_{[\kappa]}^n$ , défini par (III.60).

**Théorème III.29 (Deuxième inclusion de l'estimation sommative dans son extension)**

Soit  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première est, pour tout  $u \in \Omega$ ,  $-d\kappa(u) = a\left(\eta(u - \frac{1}{2}) - \eta(u + \frac{1}{2})\right)$ , avec  $a = \int_\Omega \max(0, -d\kappa(u))du$  et  $\eta \in \mathcal{K}(\Omega)$ . Soit  $\pi$  le noyau maxitif le plus spécifique qui domine  $\eta$ . Pour tout  $x \in \Omega$ , on a :

$$\forall \varphi \in \mathcal{D}(a, \pi), \hat{f}_\varphi^n(x) \in \underline{f}_{[\kappa]}^n(x), \quad (\text{III.73})$$

où  $\mathcal{D}(a, \pi)$  est la famille des noyaux sommatifs définie par (III.59) et  $\underline{f}_{[\kappa]}^n$  est l'intervalle d'estimation de la densité défini par (III.60).

**Preuve :** Soit  $\varphi \in \mathcal{D}(a, \pi)$ , il existe  $\xi \in \mathcal{M}(\pi)$  tel que

$$\forall u \in \Omega, -d\varphi(u) = a\left(\xi(u - \frac{1}{2}) - \xi(u + \frac{1}{2})\right).$$

Or, il existe  $\xi^x \in \mathcal{M}(\pi^x)$  tel que, d'après l'expression (II.78)

$$\forall \xi^x \in \mathcal{M}(\pi^x), \mathbb{E}_{\xi^x}(E_n^+ - E_n^-) \in \underline{\mathbb{E}}_{\pi^x}(E_n^+ - E_n^-). \quad (\text{III.74})$$

D'après l'expression (III.23), on a :

$$\forall x \in \Omega, \hat{f}_\varphi^n(x) = a \mathbb{E}_{\xi^x}(E_n^+ - E_n^-),$$

en multipliant l'expression (III.74) par  $a$ , on en déduit que

$$\forall x \in \Omega, \hat{f}_\varphi^n(x) \in a \underline{\mathbb{E}}_{\pi^x}(E_n^+ - E_n^-) = \underline{f}_{[\kappa]}^n(x).$$

□

### III.3.2 Deuxième restriction sur l'intervalle d'estimation

Dans cette deuxième restriction, nous considérons que le noyau de base  $\kappa$  est monomodale symétrique. Dans ce cas la décomposition sommative de sa dérivée première, d'après l'expression (III.25), est donnée par :

$$\forall u \in \Omega, -d\kappa(u) = a\left(\eta\left(\frac{1}{2} - u\right) - \eta\left(\frac{1}{2} + u\right)\right),$$

avec  $a = \int_\Omega \max(0, -d\kappa(u))du$  et  $\eta \in \mathcal{K}(\Omega)$ .

Soit  $\pi$  le noyau maxitif le plus spécifique qui domine  $\eta$ , on peut définir un ensemble des noyaux sommatifs, que nous notons  $\mathcal{D}^*(a, \pi)$ , par :

$$\mathcal{D}^*(a, \pi) = \left\{ \begin{array}{l} v \in \mathcal{K}_1(\Omega), \exists \xi \in \mathcal{M}(\pi), \text{ tel que} \\ \forall u \in \Omega, -dv(u) = a\left(\xi\left(\frac{1}{2} - u\right) - \xi\left(\frac{1}{2} + u\right)\right) \end{array} \right\}, \quad (\text{III.75})$$

où  $\mathcal{M}(\pi)$  est la famille des noyaux sommatifs dominés par le noyau maxitif  $\pi$ .

Cet ensemble est naturellement non vide car, par construction,  $\kappa \in \mathcal{D}^*(a, \pi)$ . On peut alors construire l'intervalle d'estimation de la densité, notée  $\underline{f}_{[\kappa]^*}^n$ , en utilisant le noyau maxitif  $\pi$  par :

$$\underline{f}_{[\kappa]^*}^n(x) = a \left( \underline{\mathbb{E}}_\pi(E_n^{x-} - E_n^{x+}) \right), \quad (\text{III.76})$$

où  $E_n^{x-}$  et  $E_n^{x+}$  sont deux fonctions de répartition empirique définies par (III.28) et  $\underline{\mathbb{E}}_\pi(\cdot)$  est l'espérance imprécise basée sur le noyau maxitif  $\pi$ .

### Algorithme de calcul des bornes

L'algorithme de calcul de l'intervalle d'estimation de la densité  $\underline{f}_{[\kappa]^*}^n$  est fortement lié au de calcul de  $\underline{f}_{[\kappa]}^n$ .

En posant, pour  $x \in \Omega$ ,  $\Delta E_n^x = E_n^{x-} - E_n^{x+}$ , l'intervalle d'estimation  $\underline{f}_{[\kappa]^*}^n$  défini par (III.76) peut s'écrire comme suit :

$$\forall x \in \Omega, \underline{f}_{[\kappa]^*}^n(x) = a \underline{\mathbb{E}}_\pi(\Delta E_n^x) = a \left[ \mathbb{C}_{N_\pi}(\Delta E_n^x), \mathbb{C}_{\Pi_\pi}(\Delta E_n^x) \right], \quad (\text{III.77})$$

où  $\mathbb{C}_{\Pi_\pi}(\Delta E_n^x)$  (rsp.  $\mathbb{C}_{N_\pi}(\Delta E_n^x)$ ) est l'intégrale de Choquet de  $\Delta E_n^x$  par rapport à la mesure de possibilité  $\Pi_\pi$  (rsp. la mesure de nécessité  $N_\pi$ ).

La fonction  $\Delta E_n^x$  est aussi une fonction en escalier définie, pour  $x \in \Omega$ , par :

$$\begin{aligned} \forall u \in \Omega, \Delta E_n^x(u) &= \frac{1}{n} \sum_{i=1}^n \left( H(x - x_i + \frac{1}{2} - u) - H(x - x_i - \frac{1}{2} + u) \right), \\ &= \frac{1}{n} \sum_{i=1}^n \begin{cases} -1 & \text{si } u \geq x - x_i + \frac{1}{2}, \\ +1 & \text{si } u \leq x_i - x + \frac{1}{2}, \\ 0 & \text{sinon.} \end{cases} \end{aligned} \quad (\text{III.78})$$

Par construction, la fonction en escalier  $\Delta E_n^x$  a  $(2n + 1)$  valeurs différentes. Soit  $\{w_i^x\}_{i \in \{0, \dots, 2n+1\}}$  l'ensemble de ces  $(2n + 2)$  valeurs, définies à partir de l'ensemble des observations  $\{x_i\}_{i \in \{1, \dots, n\}}$ , par :

$$w_i^x = \begin{cases} \frac{1}{2} + x_i - x, & \text{si } i \in \{1, \dots, n\}, \\ \frac{1}{2} + x - x_{i-n}, & \text{si } i \in \{n+1, \dots, 2n\}, \end{cases} \quad (\text{III.79})$$

où  $w_0^x = e_{min}$  et  $w_{2n+1}^x = e_{max}$ .

La fonction  $\Delta E_n^x$  change de valeur en chaque  $\{w_i^x\}_{i \in \{0, \dots, 2n+1\}}$ . Soit  $\Theta = \{0, \dots, 2n\}$  et soit  $(\cdot)$  la permutation des indices sur l'ordre des  $w_i^x$  de sorte que  $w_{(0)}^x \leq w_{(2)}^x \leq \dots \leq w_{(2n+1)}^x$ . Soit  $\{W_i^x\}_{i \in \Theta}$  les  $(2n + 1)$  intervalles définies à partir de  $w_i^x$  par :  $W_i^x = \left\{ [w_{(i)}^x, w_{(i+1)}^x[ \right\}_{i \in \Theta}$ .

Par construction, la valeur de  $\Delta E_n^x$  est constante sur chaque intervalle  $W_i^x$ . Soit  $\alpha_n^i$  la valeur constante de  $\Delta E_n^x$  sur  $W_i^x$  et soit  $c_i$  la valeur médiane de  $W_i^x$  définie par :  $c_i = \frac{w_{(i)}^x - w_{(i+1)}^x}{2}$ . Chaque valeur de  $\alpha_n^i$  peut être calculée à partir de la valeur médiane de chaque  $W_i^x$  par :

$$\alpha_n^i = \frac{1}{n} \sum_{k=1}^n \begin{cases} -1 & \text{si } c_i \geq x - x_k + \frac{1}{2}, \\ +1 & \text{si } c_i \leq x_k - x + \frac{1}{2}, \\ 0 & \text{sinon.} \end{cases} \quad (\text{III.80})$$

De même, la fonction  $\Delta E_n^x$ , définie par (III.78), peut être réécrite comme suit :

$$\forall u \in \Omega, \Delta E_n^x(u) = \sum_{i=0}^{2n} \alpha_n^i \mathbb{1}_{W_i^x}(u), \quad (\text{III.81})$$

où  $\mathbb{1}_A$  est la fonction indicatrice de  $A \subseteq \Omega$ .

L'intégrale de Choquet continue intervenant dans le calcul de  $\overline{f}_{[k]^*}^n$  (III.77), peut être remplacée par une intégrale de Choquet discrète par rapport à une distribution de possibilité associée aux  $\{W_i^x\}_{i \in \Theta}$ . D'après le Lemme III.25, on peut définir une distribution de possibilité discrète  $\tau$  sur  $\Theta$ , par :

$$\forall k \in \Theta, \tau_k = \Pi_\pi(W_k^x), \quad (\text{III.82})$$

où  $\pi$  est une distribution de possibilité continue sur  $\Omega$ .

Soit  $\Delta E_n^x$  la fonction en escalier définie par (III.78) et soit  $\alpha_n = \{\alpha_n^i\}_{i \in \Theta}$  l'ensemble des  $(2n + 1)$  valeurs de  $\Delta E_n^x$ , définis par (III.80). Soient  $(\cdot)$  la permutation des indices sur  $\{\alpha_n^i\}_{i \in \Theta}$  de sorte que  $\alpha_n^{(0)} \leq \alpha_n^{(1)} \leq \dots \alpha_n^{(2n)}$  et  $\{A_i\}_{i \in \Theta}$  les  $(2n + 1)$  sous-ensembles de  $\Theta$  définis par :  $A_{(i)} = \{(i), \dots, (2n)\}$ . L'intégrale de Choquet continue  $\mathbb{C}_{\Pi_\pi}(\Delta E_n^x)$  (rsp.  $\mathbb{C}_{N_\pi}(\Delta E_n^x)$ ) peut être calculée, d'après le théorème III.26, par une intégrale de Choquet discrètes  $\mathbb{C}_{\Pi_\tau}(\alpha_n)$  (rsp.  $\mathbb{C}_{N_\tau}(\alpha_n)$ ) de  $\alpha_n$  par rapport à la mesure de possibilité discrète  $\Pi_\tau$  (rsp. nécessité  $N_\tau$ ) sur  $\Theta$ . Ces deux intégrales de Choquet discrètes sont données par :

$$\mathbb{C}_{\Pi_\pi}(\Delta E_n^x) = \mathbb{C}_{\Pi_\tau}(\alpha_n) = \sum_{i=1}^{2n} \left( \alpha_n^{(i)} - \alpha_n^{(i-1)} \right) \Pi_\tau(A_{(i)}), \quad (\text{III.83})$$

et

$$\mathbb{C}_{N_\pi}(\Delta E_n^x) = \mathbb{C}_{N_\tau}(\alpha_n) = \sum_{i=1}^{2n} \left( \alpha_n^{(i)} - \alpha_n^{(i-1)} \right) N_\tau(A_{(i)}), \quad (\text{III.84})$$

où  $\pi$  est une distribution de possibilité continue sur  $\Omega$  et  $\tau$  est une distribution de possibilité discrète sur  $\Theta$ , définie par (III.82).

L'expression (III.84), peut être réécrite, d'après le lemme III.27, comme suit :

$$\mathbb{C}_{N_\pi}(\Delta E_n) = \mathbb{C}_{N_\tau}(\alpha_n) = \sum_{i=1}^{2n} \left( \alpha_n^{\sigma(i)} - \alpha_n^{\sigma(i-1)} \right) \Pi_\tau(A_{\sigma(i)}), \quad (\text{III.85})$$

où  $\sigma(i)$  est une permutation sur  $\{\alpha_n^i\}_{i \in \Theta}$  de sorte que  $\alpha_n^{\sigma(0)} \geq \alpha_n^{\sigma(1)} \geq \dots \alpha_n^{\sigma(2n)}$ .

L'algorithme suivant décrit plus formellement le processus de calcul de  $\{\overline{f}_{[k]^*}^n(y_j)\}_{j \in \{1, \dots, p\}}$ .

**Data:** les observations  $\{x_i\}_{i \in \{1, \dots, n\}}$ , les points  $\{y_j\}_{j \in \{1, \dots, p\}}$ , le noyau maxitif  $\pi$  et le facteur  $a$

**Result:**  $\left\{ \left[ \underline{f}_{[\kappa]^*}^n(y_j), \overline{f}_{[\kappa]^*}^n(y_j) \right] \right\}_{j \in \{1, \dots, p\}}$

```

begin
  for  $j = 1$  to  $p$  do
    – Calculer les  $(2n + 2)$  valeurs  $\{w_i^x\}_{i \in \{0, \dots, 2n+1\}}$  (par (III.79)).
    – Trier les  $\{w_i^x\}_{i \in \{0, \dots, 2n+1\}}$  par ordre croissant.
    – Calculer les  $(2n + 1)$  intervalles  $\{W_i^x\}_{i \in \{0, \dots, 2n\}}$ .
    – Calculer les  $\{\alpha_n^i\}_{i \in \{0, \dots, 2n\}}$  par l'expression (III.80).
    – Trier les  $\{\alpha_n^i\}_{i \in \{0, \dots, 2n\}}$  par ordre croissant.
    – Calculer les  $\{\tau_i\}_{i \in \{0, \dots, 2n\}}$  par l'expression (III.82).
    – Calculer  $\mathbb{C}_{\Pi_\tau}(\alpha_n)$  et  $\mathbb{C}_{N_\tau}(\alpha_n)$  :
     $\Pi \leftarrow 0$ ,  $\overline{f}_{[\kappa]^*}^n(y_j) \leftarrow 0$ 
    for  $i = 2n$  to  $1$  down to  $1$  do
       $\Pi = \max(\Pi, \mu_i)$ 
       $\overline{f}_{[\kappa]^*}^n(y_j) = \overline{f}_{[\kappa]^*}^n(y_j) + (\alpha_n^i - \alpha_n^{i-1})\Pi$ 
    end
     $\overline{f}_{[\kappa]^*}^n(y_j) = a \overline{f}_{[\kappa]^*}^n(y_j)$ 
     $\Pi \leftarrow 0$ ,  $\underline{f}_{[\kappa]^*}^n(y_j) \leftarrow 0$ 
    for  $i = 1$  to  $2n$  do
       $\Pi = \max(\Pi, \mu_i)$ 
       $\underline{f}_{[\kappa]^*}^n(y_j) = \underline{f}_{[\kappa]^*}^n(y_j) + (\alpha_n^i - \alpha_n^{i-1})\Pi$ 
    end
     $\underline{f}_{[\kappa]^*}^n(y_j) = a \underline{f}_{[\kappa]^*}^n(y_j)$ 
  end
end

```

**Algorithm 5:** Calcul de  $\left\{ \left[ \underline{f}_{[\kappa]^*}^n(y_j), \overline{f}_{[\kappa]^*}^n(y_j) \right] \right\}_{j \in \{1, \dots, p\}}$ .

### Relation avec l'estimateur de Parzen-Rosenblatt

Nous montrons dans cette partie que l'estimateur de densité de Parzen-Rosenblatt  $\widehat{f}_\kappa^n$ , défini par (I.24), est inclus dans l'intervalle d'estimation  $\underline{f}_{[\kappa]^*}^n$  défini par (III.76).

#### **Théorème III.30 (Troisième inclusion de l'estimation sommative dans son extension)**

Soit  $\kappa \in \mathcal{K}_1(\Omega)$ , dont la décomposition sommative de la dérivée première est  $\forall u \in \Omega$ ,  $-d\kappa(u) = a\left(\eta\left(\frac{1}{2} - u\right) - \eta\left(\frac{1}{2} + u\right)\right)$ , avec  $a = \int_\Omega \max(0, -d\kappa(u))du$  et  $\eta \in \mathcal{K}(\Omega)$ . Soit  $\pi$  le noyau maxitif le plus spécifique qui domine  $\eta$ . Pour tout  $x \in \Omega$ , on a :

$$\forall \varphi \in \mathcal{D}^*(a, \pi), \widehat{f}_\varphi^n(x) \in \underline{f}_{[\kappa]^*}^n(x), \quad (\text{III.86})$$

où  $\mathcal{D}^*(a, \pi)$  est la famille des noyaux sommatifs définie par (III.75) et  $\underline{f}_{[\kappa]^*}^n$  est l'intervalle d'estimation de la densité défini par (III.76).

**Preuve :** Soit  $\varphi \in \mathcal{D}^*(a, \pi)$ . Il existe  $\xi \in \mathcal{M}(\pi)$ , tel que

$$\forall u \in \Omega, -d\varphi(u) = a\left(\xi\left(\frac{1}{2} - u\right) - \xi\left(\frac{1}{2} + u\right)\right).$$

Or, d'après l'expression (II.78), on a :

$$\forall \xi \in \mathcal{M}(\pi), \mathbb{E}_\xi(E_n^{x^-} - E_n^{x^+}) \in \underline{\mathbb{E}}_\pi(E_n^{x^-} - E_n^{x^+}). \quad (\text{III.87})$$

D'après l'expression (III.29), on a :

$$\forall x \in \Omega, \hat{f}_\phi^n(x) = a \mathbb{E}_\xi(E_n^{x^-} - E_n^{x^+}),$$

en multipliant l'expression (III.87) par  $a$ , on en déduit que

$$\forall x \in \Omega, \hat{f}_\phi^n(x) \in a \underline{\mathbb{E}}_\pi(E_n^{x^-} - E_n^{x^+}) = \underline{f}_{[\kappa]^*}^n(x).$$

□

# Chapitre IV

## Expérimentations

### Introduction

Nous avons présenté, dans le chapitre III, une proposition de méthodes intervallistes d'estimation de la densité de probabilité sous-jacente à un ensemble fini d'observations. Nous avons donné un certain nombre de propriétés intéressantes. Dans ce chapitre, nous proposons d'illustrer ces propriétés théoriques par des expérimentations.

La section IV.1 montre, avec des exemples numériques, les techniques de calcul que nous employons pour dominer la dérivée première des noyaux sommatifs que nous utilisons.

La section IV.2 présente cinq expérimentations illustrant les propriétés théoriques de l'intervalle d'estimation de la densité présentées au chapitre III. Ces expérimentations s'appuient sur des données simulées nous permettant ainsi de comparer les résultats obtenus à une vérité terrain.

Pour illustrer nos expérimentations, nous utilisons deux noyaux sommatifs. Nous utilisons d'une part le noyau d'*Epanechnikov* qui est un noyau de référence en estimation non-paramétrique de la densité. Ce noyau est connu pour minimiser le critère AMISE en estimation de densité (cf. section I.2.3.2). Nous l'avons ainsi choisi parce qu'il a un comportement singulier dans l'approche d'estimation que nous proposons (voir section III.2.3). Nous utilisons d'autre part ainsi qu'un noyau, ayant des propriétés intéressantes dues à sa symétrie, que nous nommerons noyau *cosinus symétrisé* dans la suite de ce document. Cette propriété de symétrie nous permet de comparer les trois décompositions présentées en section III.1.3.

Nous basons nos expérimentations sur des observations simulées d'une variable aléatoire dont la densité de probabilité est une distribution bimodale obtenue en contaminant une loi normale de moyenne 3 et de variance 1 par une loi normale de moyenne 8 et de variance 4.

Pour simplifier la lecture de ces expérimentations, nous bloquons un certain nombre de notations. Ainsi  $\kappa$  représente le *noyau sommatif de référence*,  $(\eta^-, \eta^+)$  les *deux noyaux sommatifs de la décomposition de sa dérivée première*,  $(\pi^-, \pi^+)$  les *deux noyaux maxitifs obtenus par la transformation probabilité-possibilité objective (domination plus spécifique)* et  $(\pi^-, \pi^+)$  les *deux noyaux maxitifs obtenus par la transformation probabilité-possibilité subjective (domination moins spécifique)*.  $\eta$  représente le *noyau sommatif de translation* et  $\pi$  le *noyau maxitif le plus spécifique* qui le domine. Il est important de souligner que nous ne donnons le calcul de  $(\pi^-, \pi^+)$  que pour le noyau d'Epanechnikov en raison de son



comportement particulier dans notre approche d'estimation.

## IV.1 Décomposition sommative d'un noyau et domination de cette décomposition

Dans cette section, nous présentons trois dominations (cf. section III.1.3). La première domination est générale (cf. section III.1.3). La seconde ne concerne que les noyaux symétriques (cf. section III.1.4.3). Enfin la dernière nécessite que les deux noyaux  $\eta^+$  et  $\eta^-$ , intervenant dans la décomposition puissent se déduire l'un de l'autre par simple translation (cf. section III.1.4.2). Les deux premières méthodes peuvent être appliquées sur les deux noyaux (noyau d'Epanechnikov et noyau cosinus symétrisé). La dernière méthode ne peut être appliquée que sur le noyau cosinus symétrisé.

### IV.1.1 Cas où le noyau est quelconque

Quand le noyau de référence  $\kappa$  est quelconque (mais sommatif), la décomposition sommative de sa dérivée première est donnée (III.11) par :

$$\forall u \in \Omega, -d\kappa(u) = a \left( \eta^+(u) - \eta^-(u) \right).$$

**Exemple IV.1 (Noyau d'Epanechnikov)** *Nous considérons, dans cet exemple, le noyau d'Epanechnikov comme noyau de référence (III.2.3) :*

$$\kappa(u) = \frac{3}{4}(1 - u^2) \mathbb{1}_{[-1,1]}(u).$$

*Les deux noyaux sommatifs  $\eta^-$  et  $\eta^+$  intervenant dans la décomposition de sa dérivée sont donnés par :*

$$\eta^-(u) = -2u \mathbb{1}_{[-1,0]}(u), \tag{IV.1}$$

*et*

$$\eta^+(u) = 2u \mathbb{1}_{[0,1]}(u). \tag{IV.2}$$

*Les deux noyaux maxitifs  $\pi^-$  et  $\pi^+$  sont donnés par :*

$$\pi^-(u) = u^2 \mathbb{1}_{[-1,0]}(u), \tag{IV.3}$$

*et*

$$\pi^+(u) = u^2 \mathbb{1}_{[0,1]}(u). \tag{IV.4}$$

*Les deux noyaux maxitifs  $\pi^-$  et  $\pi^+$  sont donnés par :*

$$\pi^-(u) = (-2u - u^2) \mathbb{1}_{[-1,0]}(u), \tag{IV.5}$$

*et*

$$\pi^+(u) = (2u - u^2) \mathbb{1}_{[0,1]}(u). \tag{IV.6}$$

□

Nous détaillons maintenant comment nous obtenons ces résultats.

**Calcul IV.2.** La décomposition de la dérivée première de  $\kappa$  est donnée par :

$$\begin{aligned} -d\kappa(u) &= \frac{3}{2} |u| \mathbb{1}_{[-1,1]}(u) \\ &= \frac{3}{2} u \mathbb{1}_{[0,1]}(u) - \frac{3}{2} u \mathbb{1}_{[-1,0]}(u) \\ &= d\kappa^+(u) - d\kappa^-(u), \end{aligned}$$

d'après la Propriété III.11, on a :

$$\begin{aligned} a &= \int_{\Omega} \max(0, -d\kappa(u)) du, \text{ par la symétrie de } \kappa \\ &= \int_{\Omega} d\kappa^+(u) du \\ &= \frac{3}{2} \int_0^1 u du \\ &= \frac{3}{4}, \end{aligned}$$

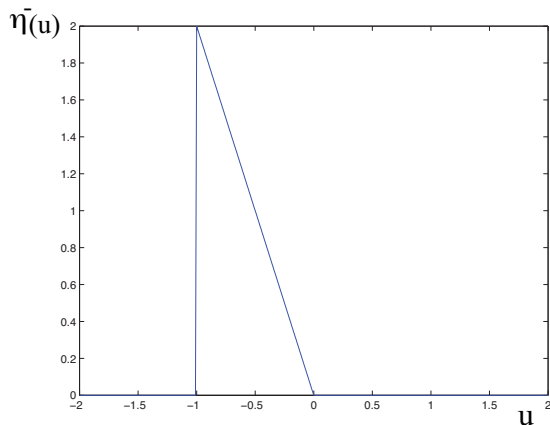
or,  $\eta^+(u) = \frac{d\kappa^+}{a}$  (voir Propriété III.11), d'où

$$\eta^+(u) = 2u\mathbb{1}_{[0,1]}(u).$$

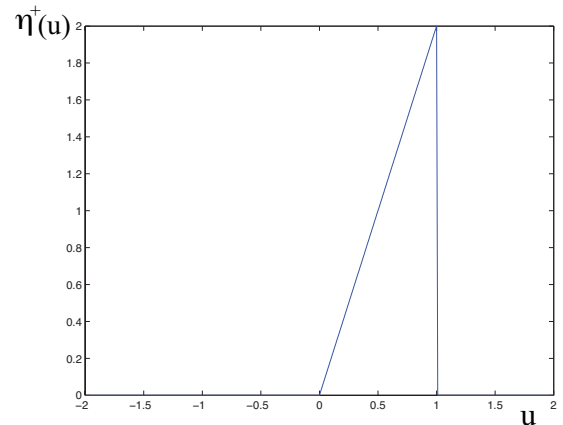
De la même manière on obtient

$$\eta^-(u) = -2u\mathbb{1}_{[-1,0]}(u).$$

La Figure IV.1 (a) (resp. (b)) illustre le noyau sommatif  $\eta^-$  (IV.1) (resp.  $\eta^+$  (IV.2)).



(a)  $\eta^-(u) = -2u\mathbb{1}_{[-1,0]}(u)$ .



(b)  $\eta^+(u) = 2u\mathbb{1}_{[0,1]}(u)$ .

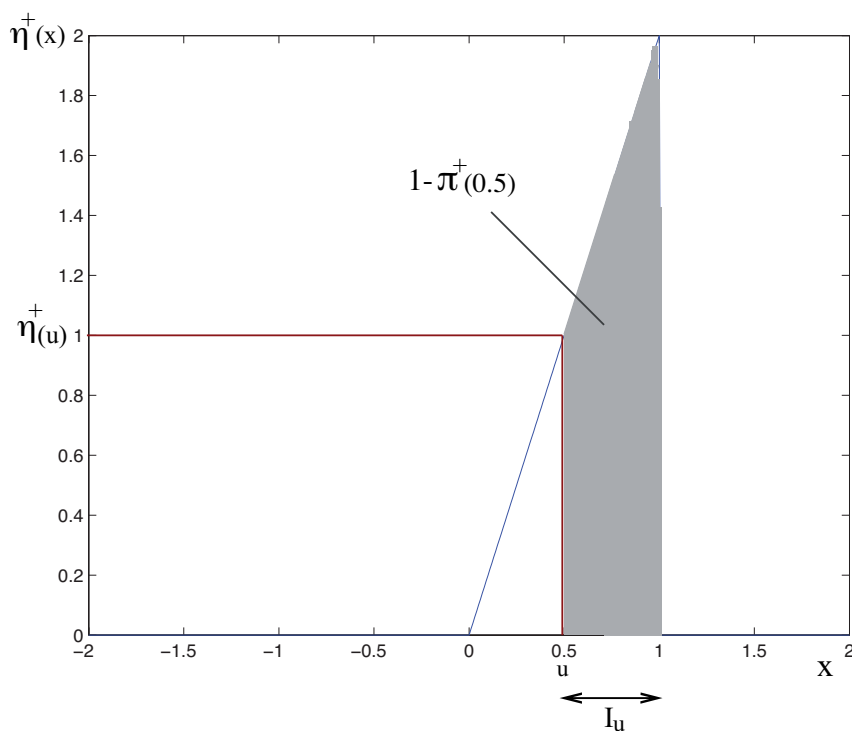
**Figure IV.1 :** Les deux noyaux sommatifs  $\eta^-$  (a) et  $\eta^+$  (b) intervenant dans la décomposition sommative du noyau d'Epanechnikov.

Le noyau maxitif  $\pi^+$  est donné, d'après l'expression (II.69), par :

$$\begin{aligned} \forall u \in [0, 1], \pi^+(u) &= 1 - P_{\eta^+}(I_u) \\ &= 1 - \int_u^1 \eta^+(v) dv \\ &= 1 - \int_u^1 2v dv \\ &= u^2. \end{aligned}$$

avec  $I_u = \{x \in \Omega : \eta^+(x) \geq \eta^+(u)\}$ .

La Figure IV.2 illustre la technique de calcul du noyau maxitif  $\pi^+$  (IV.4). On peut voir sur la Figure IV.2, le noyau sommatif  $\eta^+$  (IV.2) (en bleu), la probabilité  $P_{\eta^+}(I_u) = 1 - \pi^+(u)$  (en gris) et l'intervalle de confiance  $I_u$ .



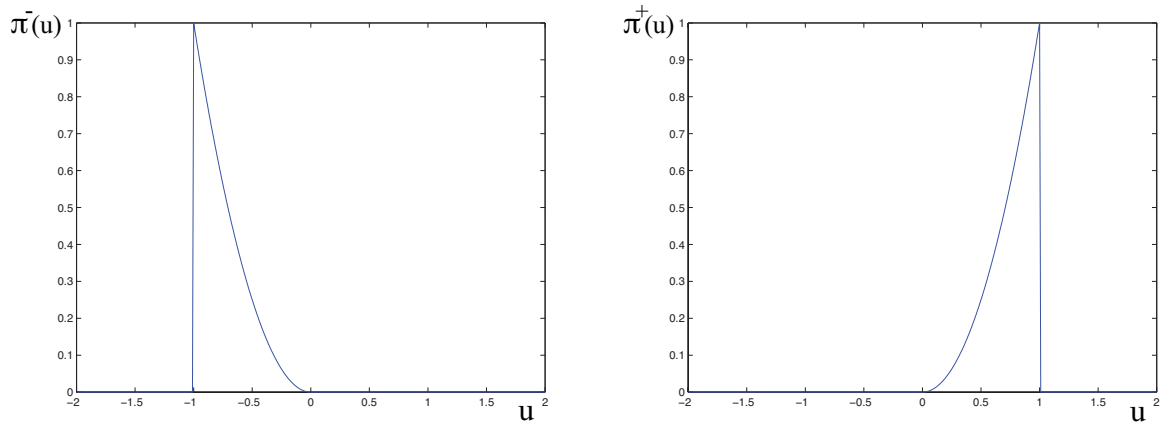
**Figure IV.2 :** Le noyau sommatif  $\eta^+(x) = 2x\mathbb{1}_{[0,1]}(x)$  et l'intervalle de confiance  $I_{u=0.5}$ .

De la même manière, on obtient :

$$\begin{aligned} \forall u \in [-1, 0], \pi^-(u) &= 1 - P_{\eta^-}(I_u) \\ &= 1 - \int_{-1}^u \eta^-(v) dv \\ &= 1 + \int_{-1}^u 2v dv \\ &= u^2. \end{aligned}$$

L'illustration de la technique de calcul de  $\pi^-$  se fait de la même manière que celle de  $\pi^+$ .

La Figure IV.3 (a) (resp. (b)) illustre (en bleu) le noyau maxitif  $\pi^-$  (IV.3) (resp.  $\pi^+$  (IV.4)).



(a)  $\pi^-(u) = u^2 \mathbb{1}_{[-1,0]}(u)$ .

(b)  $\pi^+(u) = u^2 \mathbb{1}_{[0,1]}(u)$ .

**Figure IV.3 :** Les deux noyaux maxitifs  $\pi^-$  (a) et  $\pi^+$  (b) intervenant dans la domination objective de la décomposition sommative du noyau d'Epanechnikov.

Le noyau maxitif  $\pi^+$  est donné, d'après l'expression (II.72), par :

$$\begin{aligned} \pi^+(u) &= \pi^+(u) + \lambda(I_u)\eta^+(u) \\ &= u^2 + (1 - u)(2u) \\ &= 2u - u^2, \end{aligned}$$

$\lambda(I_u)$  étant la mesure de Lebesgue<sup>1</sup> de l'intervalle  $I_u$ .

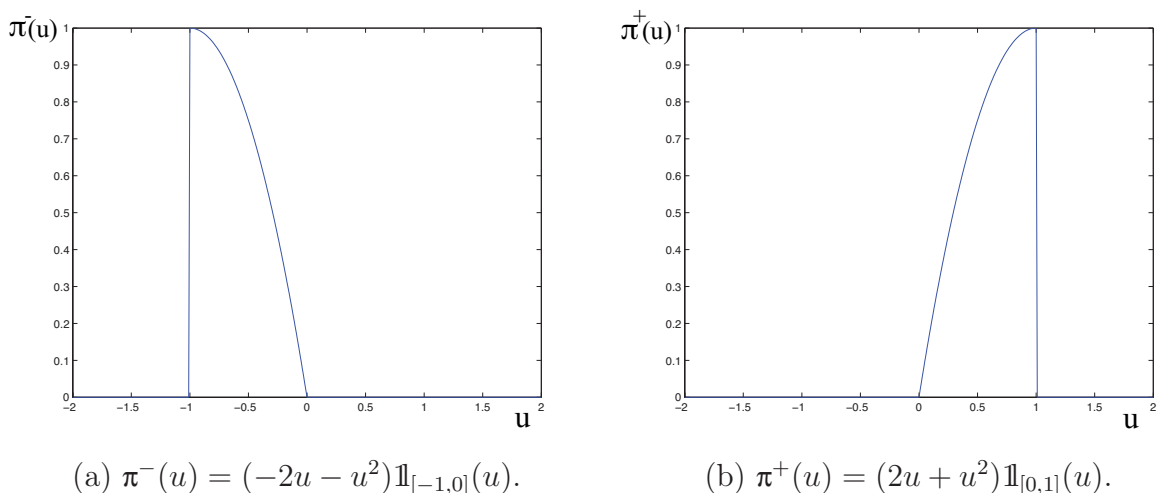
Le noyau maxitif  $\pi^-$  est donné par :

$$\begin{aligned} \pi^-(u) &= \pi^-(u) + \lambda(I_u)\eta^-(u) \\ &= u^2 - (1 + u)(2u) \\ &= -2u - u^2. \end{aligned}$$

La Figure IV.4 (a) (resp. (b)) illustre le noyau maxitif  $\pi^-$  (resp.  $\pi^+$ ).

---

1. Soit  $A = [a_1, a_2]$  un intervalle de  $\mathbb{R}$ , la mesure de Lebesgue de  $A$ , notée  $\lambda(A)$ , est donnée par :  $\lambda(A) = |a_2 - a_1|$  ([4] page 17).



**Figure IV.4 :** Les deux noyaux maxitifs  $\pi^-$  (a) et  $\pi^+$  (b) intervenant dans la domination subjective de la décomposition sommative du noyau d'Epanechnikov.

□

**Exemple IV.3 (Noyau Cosinus Symétrisé)** *Ce noyau est construit en se basant sur la fonction cosinus de façon à ce que les deux noyaux intervenant dans la décomposition sommative de sa dérivée première se déduisent l'un de l'autre par translation. Ce noyau est défini par :*

$$\kappa(u) = \frac{1}{2}(1 + \cos(u\pi))\mathbb{1}_{[-1,1]}(u). \quad (\text{IV.7})$$

*Les deux noyaux sommatifs  $\eta^-$  et  $\eta^+$  intervenant dans la décomposition de sa dérivée sont donnés par :*

$$\eta^-(u) = -\frac{\pi}{2}\sin(u\pi)\mathbb{1}_{[-1,0]}(u), \quad (\text{IV.8})$$

et

$$\eta^+(u) = \frac{\pi}{2}\sin(u\pi)\mathbb{1}_{[0,1]}(u). \quad (\text{IV.9})$$

*Les deux noyaux maxitifs  $\pi^-$  et  $\pi^+$  sont donnés par :*

$$\pi^-(u) = \begin{cases} 1 + \cos(u\pi) & \text{si } u \in [-1, -\frac{1}{2}], \\ 1 - \cos(u\pi) & \text{si } u \in [-\frac{1}{2}, 0], \end{cases} \quad (\text{IV.10})$$

et

$$\pi^+(u) = \begin{cases} 1 - \cos(u\pi) & \text{si } u \in [0, \frac{1}{2}], \\ 1 + \cos(u\pi) & \text{si } u \in [\frac{1}{2}, 1]. \end{cases} \quad (\text{IV.11})$$

Nous détaillons maintenant comment nous obtenons ces résultats.

**Calcul IV.4.** La décomposition de la dérivée première de  $\kappa$  (IV.7) est donnée par :

$$\begin{aligned} -d\kappa(u) &= \frac{\pi}{2} \sin(|u| \pi) \mathbb{1}_{[-1,1]}(u) \\ &= \frac{\pi}{2} \sin(u\pi) \mathbb{1}_{[0,1]}(u) + \frac{\pi}{2} \sin(u\pi) \mathbb{1}_{[-1,0]}(u) \\ &= d\kappa^+(u) - d\kappa^-(u), \end{aligned}$$

on a :

$$\begin{aligned} a &= \int_0^1 d\kappa^+(u) du \\ &= \frac{\pi}{2} \int_0^1 \sin(u\pi) du \\ &= 1, \end{aligned}$$

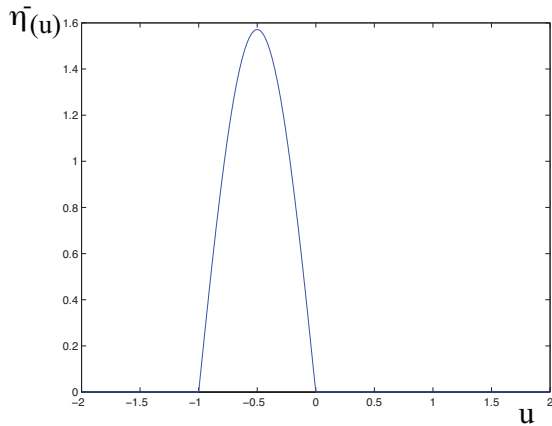
or,  $\eta^+(u) = \frac{d\kappa^+}{a}$ , d'où

$$\eta^+(u) = \frac{\pi}{2} \sin(u\pi) \mathbb{1}_{[0,1]}(u).$$

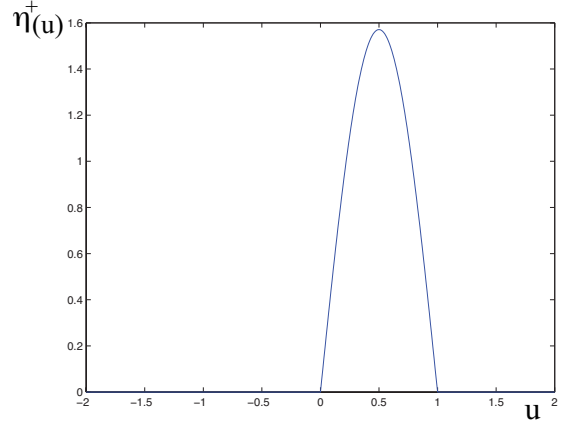
De la même manière on obtient

$$\eta^-(u) = -\frac{\pi}{2} \sin(u\pi) \mathbb{1}_{[-1,0]}(u).$$

La Figure IV.5 (a) (resp. (b)) illustre le noyau sommatif  $\eta^-$  (resp.  $\eta^+$ ).



(a)  $\eta^-(u) = -\frac{\pi}{2} \sin(u\pi) \mathbb{1}_{[-1,0]}(u)$ .



(b)  $\eta^+(u) = \frac{\pi}{2} \sin(u\pi) \mathbb{1}_{[0,1]}(u)$ .

**Figure IV.5 :** Les deux noyaux sommatifs  $\eta^-$  (a) et  $\eta^+$  (b) intervenant dans la décomposition sommative du noyau cosinus symétrisé.

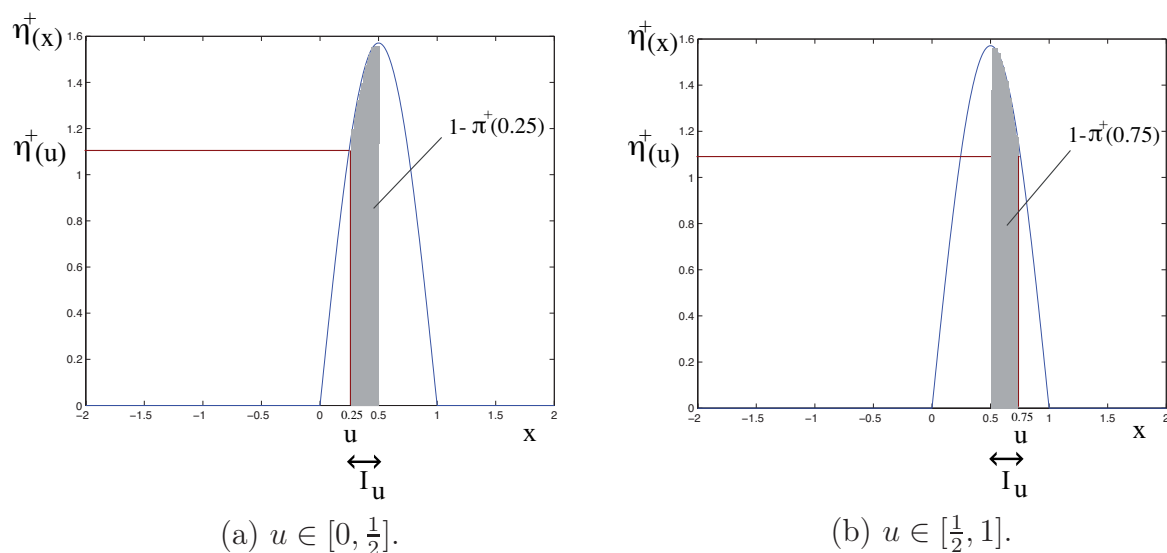
Le noyau maxitif  $\pi^+$  est donné, pour  $u \in [0, \frac{1}{2}]$ , par :

$$\begin{aligned} \pi^+(u) &= 1 - P_{\eta^+}(I_u) \\ &= 1 - 2 \int_u^{\frac{1}{2}} \frac{\pi}{2} \sin(v\pi) dv \\ &= 1 - \cos(u\pi), \end{aligned}$$

et pour  $u \in [\frac{1}{2}, 1]$  :

$$\begin{aligned}\pi^+(u) &= 1 - P_{\eta^+}(I_u) \\ &= 1 - 2 \int_{\frac{1}{2}}^u \frac{\pi}{2} \sin(v\pi) dv \\ &= 1 + \cos(u\pi).\end{aligned}$$

La Figure IV.6 (a) (resp. (b)) illustre la technique de calcul de noyau maxitif  $\pi^+$  (IV.11) pour  $u \in [0, \frac{1}{2}]$  (resp.  $u \in [\frac{1}{2}, 1]$ ). On peut voir sur la Figure IV.6 le noyau sommatif  $\eta^+$  (IV.9) (en bleu), la probabilité  $P_{\eta^+}(I_u) = 1 - \pi^+(u)$  (en gris) et l'intervalle de confiance  $I_u$ .

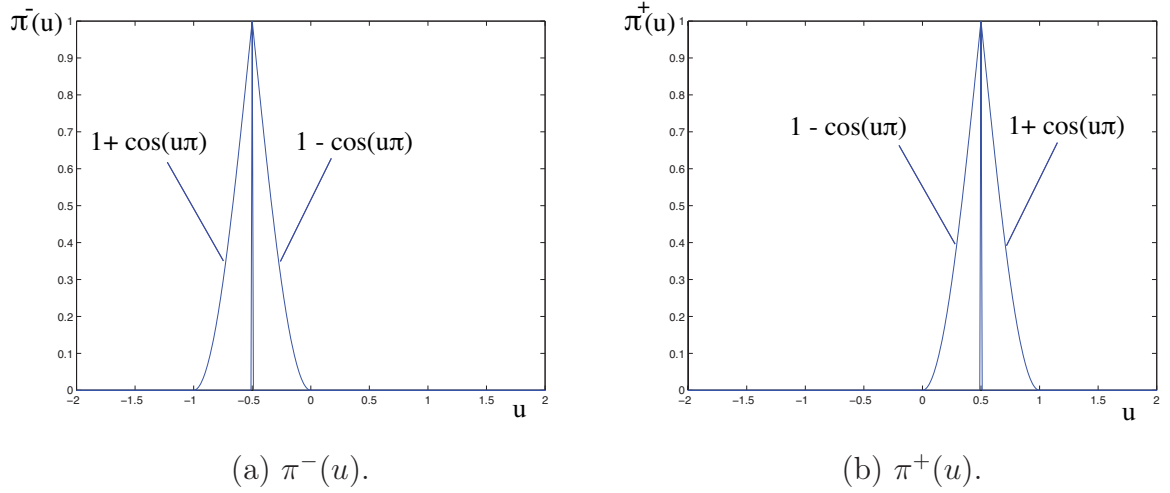


**Figure IV.6** : Le noyau sommatif  $\eta^+(x) = \frac{\pi}{2} \sin(x\pi) \mathbb{1}_{[0,1]}(x)$  et l'intervalle de confiance  $I_{u=0.25}$  (resp.  $I_{u=0.75}$ ) (a) (resp. (b)).

De la même manière, on obtient :

$$\pi^-(u) = \begin{cases} 1 + \cos(u\pi) & \text{si } u \in [-1, -\frac{1}{2}], \\ 1 - \cos(u\pi) & \text{si } u \in [-\frac{1}{2}, 0]. \end{cases}$$

La Figure IV.7 (a) (resp. (b)) illustre le noyau maxitif  $\pi^-$  (resp.  $\pi^+$ ).



**Figure IV.7 :** Les deux noyaux maxitifs  $\pi^-$  (a) et  $\pi^+$  (b) intervenant dans la domination objective de la décomposition sommative du noyau cosinus symétrisé.

□

### IV.1.2 Cas où le noyau est simplement symétrique

Dans le cas où le noyau sommatif  $\kappa$  est symétrique, la décomposition sommative de sa dérivée première est donnée par (III.25) :

$$\forall u \in \Omega, -d\kappa(u) = a\left(\eta\left(\frac{1}{2} - u\right) - \eta\left(\frac{1}{2} + u\right)\right).$$

**Exemple IV.5.** Dans cet exemple, nous considérons le noyau d'Epanechnikov (III.2.3) comme noyau de référence.

Le noyau sommatif de translation  $\eta$  est donné par :

$$\eta(u) = (1 - 2u)\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u). \quad (\text{IV.12})$$

Le noyau maxitif de domination  $\pi$  est donné par :

$$\pi(u) = \left(\frac{1}{4} - u + u^2\right)\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u). \quad (\text{IV.13})$$

Nous détaillons maintenant comment nous obtenons ce résultat.

**Calcul IV.6.** Soit  $\eta^+$  le noyau sommatif défini par (IV.2), le noyau sommatif de translation  $\eta$  (IV.1) peut être obtenu par un simple centrage du mode de  $\eta^+$  :

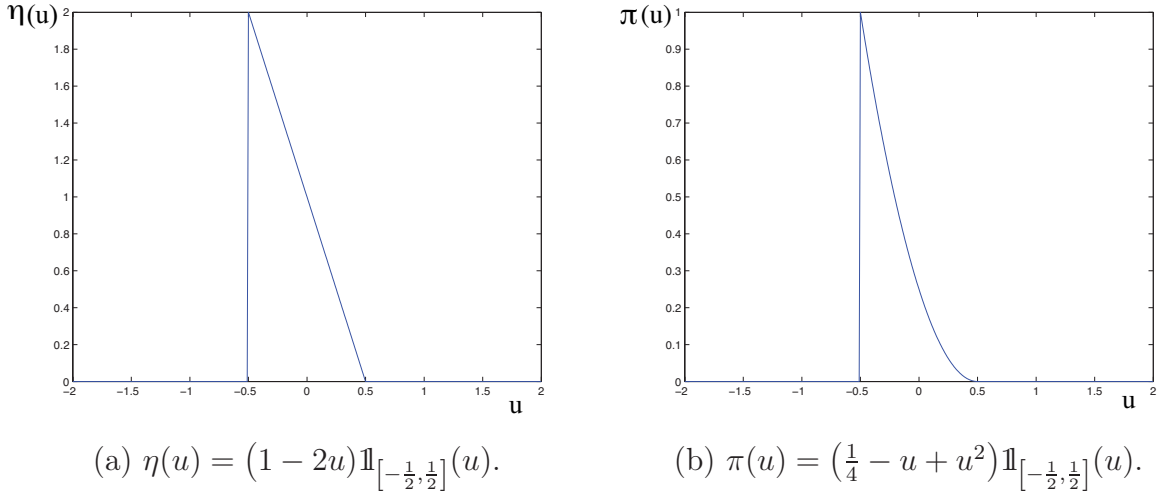
$$\begin{aligned} \eta(u) &= \eta^+\left(\frac{1}{2} - u\right) \\ &= 2\left(\frac{1}{2} - u\right)\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u) \\ &= (1 - 2u)\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u). \end{aligned}$$



De la même façon, le noyau maxitif de domination  $\pi$  peut être obtenu par un simple centrage du noyau maxitif  $\pi^+$  défini par (IV.3) :

$$\begin{aligned}\pi(u) &= \pi^+\left(\frac{1}{2} - u\right) \\ &= \left(\frac{1}{2} - u\right)^2 \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u) \\ &= \left(\frac{1}{4} - u + u^2\right) \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u).\end{aligned}$$

On peut aussi calculer  $\pi$  via le noyau sommatif  $\eta$  en utilisant la méthode de domination de noyau maxitif (voir Calcul IV.2). La Figure IV.8 (a) (resp. (b)) illustre le noyau sommatif  $\eta$  (IV.12) (resp. maxitif  $\pi$  (IV.13) ).



**Figure IV.8 :** Le noyau sommatif de translation  $\eta$  (a) intervenant dans la décomposition sommative du noyau d'Epanechnikov et le noyau maxitif  $\pi$  (b) dominant, via une domination objective, le noyau  $\eta$ .

□

**Exemple IV.7.** Dans cet exemple, nous considérons le noyau cosinus symétrisé défini par (IV.7) comme noyau de référence.

Le noyau sommatif de translation  $\eta$  est donné par :

$$\eta(u) = \frac{\pi}{2} \cos(u\pi) \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u). \quad (\text{IV.14})$$

Le noyau maxitif  $\pi$  dominant  $\eta$  est donné par :

$$\pi(u) = (1 - \sin(|u|\pi)) \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u). \quad (\text{IV.15})$$

**Calcul IV.8.** Soit  $\eta^+$  le noyau sommatif défini par (IV.9), le noyau sommatif  $\eta$  (IV.14)

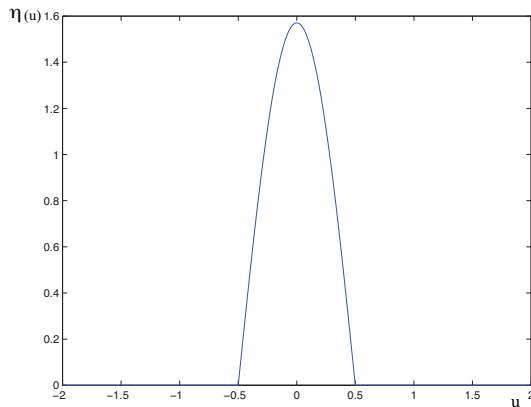
peut être calculé par la translation suivante :

$$\begin{aligned}\eta(u) &= \eta^+\left(\frac{1}{2} - u\right) \\ &= \frac{\pi}{2} \sin\left(\left(\frac{1}{2} - u\right)\pi\right) \mathbb{1}_{[0,1]}(u) \\ &= \frac{\pi}{2} \cos(u\pi) \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u).\end{aligned}$$

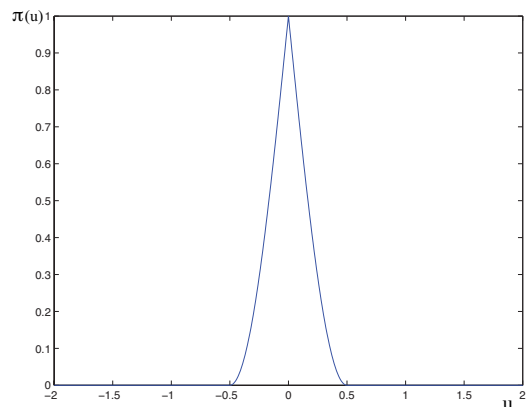
Soit  $\pi^+$  le noyau maxitif défini dans (IV.11), le noyau maxitif  $\pi$  peut être calculé :

$$\begin{aligned}\pi(u) &= \pi^+\left(\frac{1}{2} - u\right) \\ &= \begin{cases} 1 - \cos\left(\left(\frac{1}{2} - u\right)\pi\right) & \text{si } u \in \left[0, \frac{1}{2}\right] \\ 1 + \cos\left(\left(\frac{1}{2} - u\right)\pi\right) & \text{si } u \in \left[\frac{1}{2}, 1\right] \end{cases} \\ &= \begin{cases} 1 - \sin(u\pi) & \text{si } u \in \left[0, \frac{1}{2}\right] \\ 1 + \sin(u\pi) & \text{si } u \in \left[-\frac{1}{2}, 0\right] \end{cases} \\ &= (1 - \sin(|u| \pi)) \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u).\end{aligned}$$

La Figure IV.9 (a) (resp. (b)) illustre le noyau sommatif  $\eta$  (IV.14) (resp. maxitif  $\pi$  (IV.15))



(a)  $\eta(u) = \frac{\pi}{2} \cos(u\pi) \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u)$ .



(b)  $\pi(u) = (1 - \sin(|u| \pi)) \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u)$ .

**Figure IV.9 :** Le noyau sommatif de translation  $\eta$  (a) intervenant dans la décomposition sommative du noyau cosinus symétrisé et le noyau maxitif  $\pi$  (b) dominant, via une domination objective, le noyau  $\eta$

□

### IV.1.3 Cas où les deux noyaux de la décomposition peuvent être déduits l'un de l'autre par simple translation.

Seul le noyau cosinus symétrisé (IV.7) vérifie cette particularité. Dans ce cas, le noyau sommatif de translation ainsi le noyau maxitif de domination sont les mêmes que dans la

section IV.1.2 (voir les expressions (IV.14) et (IV.15)). Nous proposons ici une autre façon de calculer ces noyaux. Ce calcul est basé sur la décomposition de la dérivée première de  $\kappa$  donnée (III.27) :

$$\forall u \in \Omega, -d\kappa(u) = a\left(\eta\left(u - \frac{1}{2}\right) - \eta\left(u + \frac{1}{2}\right)\right).$$

Le noyau sommatif de translation  $\eta$  est donné par (IV.14) :

$$\eta(u) = \frac{\pi}{2} \cos(u\pi) \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u).$$

Le noyau maxitif de domination  $\pi$  est donné par (IV.15) :

$$\pi(u) = \left(1 - \sin(|u|\pi)\right) \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u).$$

**Calcul IV.9.** Soit  $\eta^+$  le noyau sommatif défini par (IV.9), le noyau sommatif de translation  $\eta$  (IV.14) peut être recalculé par la translation suivante :

$$\begin{aligned} \eta(u) &= \eta^+\left(u + \frac{1}{2}\right) \\ &= \frac{\pi}{2} \sin\left(\left(u + \frac{1}{2}\right)\pi\right) \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u) \\ &= \frac{\pi}{2} \cos(u\pi) \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u). \end{aligned}$$

Soit  $\pi^+$  le noyau maxitif défini par (IV.11), le noyau maxitif  $\pi$  (IV.15) peut être recalculé :

$$\begin{aligned} \pi(u) &= \pi^+\left(u + \frac{1}{2}\right) \\ &= \begin{cases} 1 - \cos\left(\left(u + \frac{1}{2}\right)\pi\right) & \text{si } u \in [0, \frac{1}{2}] \\ 1 + \cos\left(\left(u + \frac{1}{2}\right)\pi\right) & \text{si } u \in [\frac{1}{2}, 1] \end{cases} \\ &= \begin{cases} 1 + \sin(u\pi) & \text{si } u \in [-\frac{1}{2}, 0] \\ 1 - \sin(u\pi) & \text{si } u \in [0, \frac{1}{2}] \end{cases} \\ &= \left(1 - \sin(|u|\pi)\right) \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(u). \end{aligned}$$

□

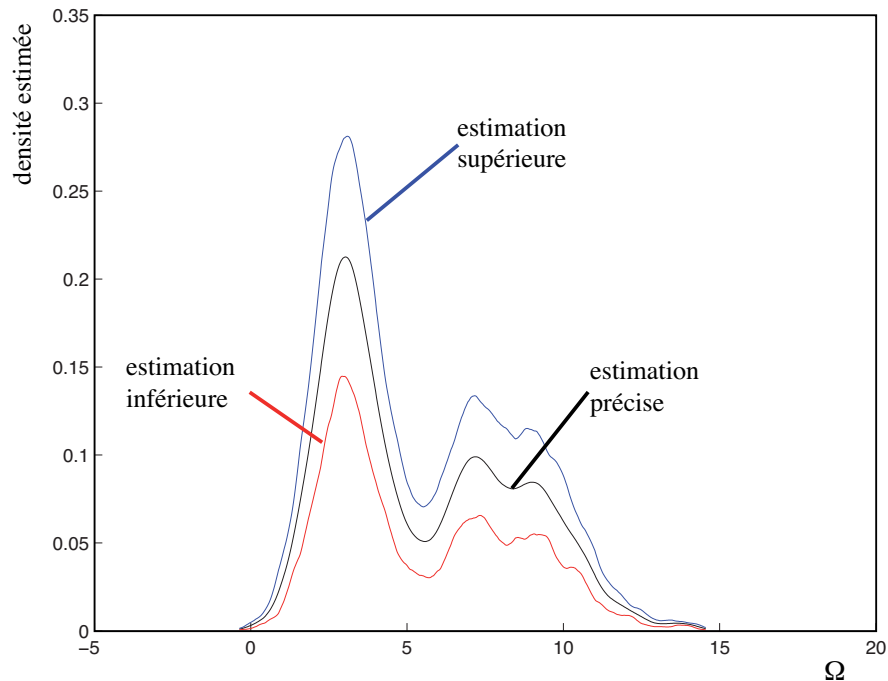
## IV.2 Etude expérimentale de l'estimation imprécise de la densité

Dans la section précédente, nous avons présenté les techniques permettant de dominer la décomposition sommative d'un noyau de référence. Nous avons appliqué ces techniques à deux noyaux : celui d'Epanechnikov et celui que nous nommons cosinus symétrisé. Dans cette section, nous nous servons de ces calculs pour mettre en évidence les propriétés de l'estimateur de densité que nous proposons et que nous avons présentées au chapitre III. Ces expérimentations sont basées sur les données simulées présentées en introduction.

### IV.2.1 Estimation précise et imprécise de la densité : cas général

Nous proposons, dans cette partie, deux expériences illustrant le Théorème III.18 et la Propriété III.19. Le Théorème III.18 dit que l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_{\kappa_\Delta}^n$  (I.24) estimé avec un noyau  $\kappa_\Delta$  est inclus dans l'intervalle d'estimation de densité  $\underline{\bar{f}}_{(\kappa_\Delta)}^n$  défini par l'expressions (III.31). La propriété III.19 dit que l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_{\kappa_\Delta}^n$  estimé avec le noyau d'Epanechnikov défini par l'expression (III.2.3) n'est autre que l'estimation inférieure  $\underline{f}_{(\kappa_\Delta)}^n$  de  $\underline{\bar{f}}_{(\kappa_\Delta)}^n$ . Pour ces deux expériences, nous avons simulé 1000 observations  $\{x_i\}_{i \in \{1, \dots, 1000\}}$  issues de la loi bimodale que nous considérons tout au long de ce chapitre. L'intervalle de référence  $\Omega = [-5, 20]$  est subdivisé en 500 échantillons régulièrement espacés  $\{y_i\}_{i \in \{1, \dots, 500\}}$ .

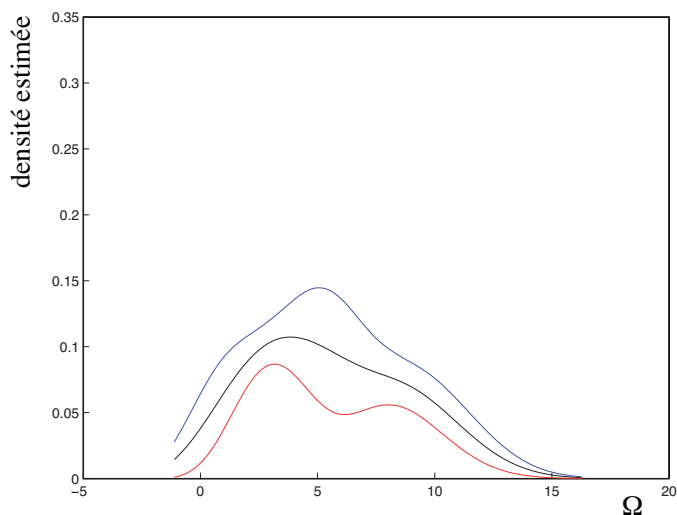
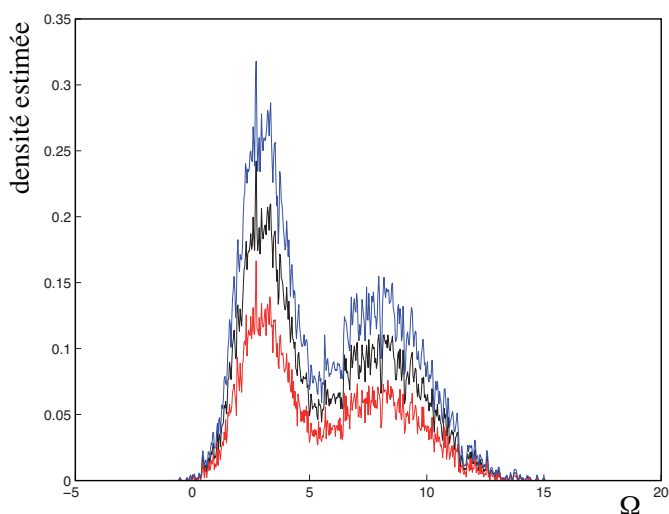
Dans la première expérience, nous avons réalisé l'estimation précise  $\hat{f}_{\kappa_\Delta}^n$  en utilisant le noyau cosinus symétrisé défini par l'expression (IV.7). La largeur de bande  $\Delta = 1$  semble la plus adaptée à l'estimation de cette densité de probabilité pour ce nombre d'observations. Nous avons réalisé l'estimation imprécise de densité  $\underline{\bar{f}}_{(\kappa_\Delta)}^n$  en utilisant le couple de noyaux maxitifs, dominant la décomposition sommative de  $\kappa_\Delta$ , le plus spécifique  $(\pi_\Delta^-, \pi_\Delta^+)$  défini par les expressions (IV.10) et (IV.11). Sur la Figure IV.10, l'estimation précise  $\hat{f}_{\kappa_\Delta}^n$  est dessinée en noir, la borne inférieure de l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$  est en rouge et sa borne supérieure en bleu. La superposition des deux estimations montre que, pour chaque valeur  $\{y_i\}_{i \in \{1, \dots, 500\}}$  de l'intervalle de référence  $\Omega$ ,  $\hat{f}_{\kappa_\Delta}^n(y_i) \in \underline{\bar{f}}_{(\kappa_\Delta)}^n(y_i)$ .



**Figure IV.10 :** Superposition de l'estimation précise (noyau cosinus symétrisé) et l'estimation imprécise (domination plus spécifique).

Cette propriété est indépendante du choix de la largeur de bande  $\Delta$ . Pour illustrer cette indépendance, la Figure IV.11 propose la même expérience, en utilisant le même

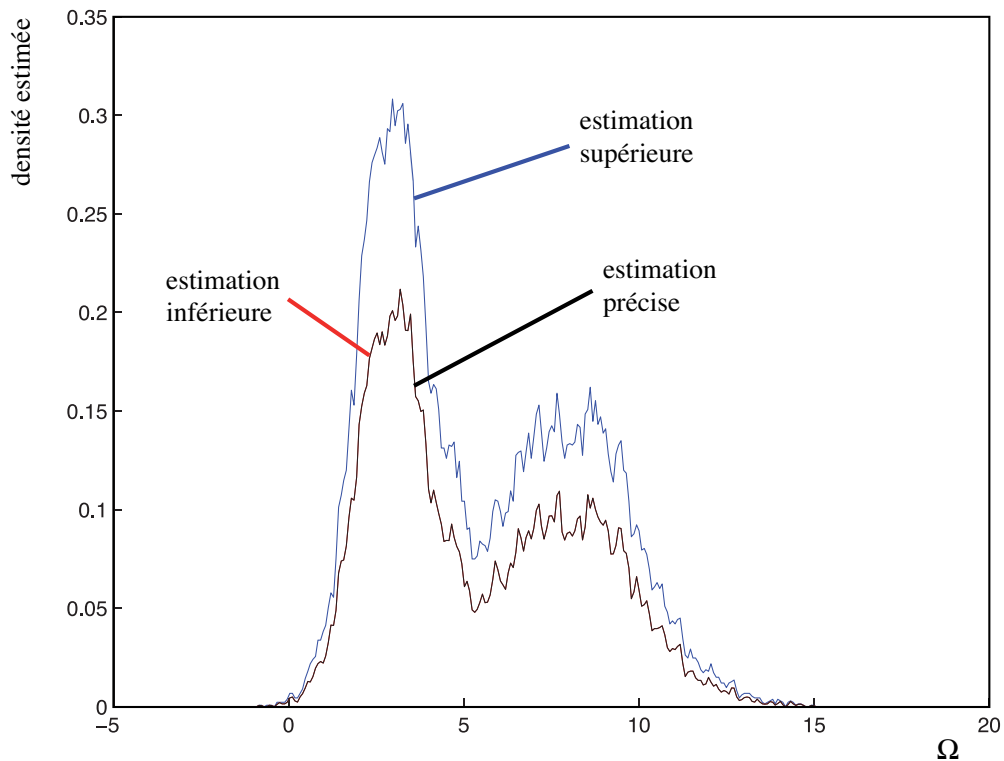
noyau sommatif  $\kappa_\Delta$  (IV.7) avec deux valeurs différentes de largeur de bande :  $\Delta = 5$  (Figure IV.11 (a)) et  $\Delta = 0.05$  (Figure IV.11 (b)). Pour  $\Delta = 5$ , on obtient une estimation trop lissée (la bimodalité de la distribution a presque disparu) tandis que pour  $\Delta = 0.05$  la densité de probabilité est mal estimée car très bruitée (difficile à interpréter).

(a)  $\Delta = 5$ (b)  $\Delta = 0.05$ 

**Figure IV.11 :** Superposition de l'estimation imprécise et l'estimation précise pour  $\Delta = 5$  (a) et  $\Delta = 0.05$  (b). Le code de couleurs utilisé est le même que dans la Figure. IV.10.

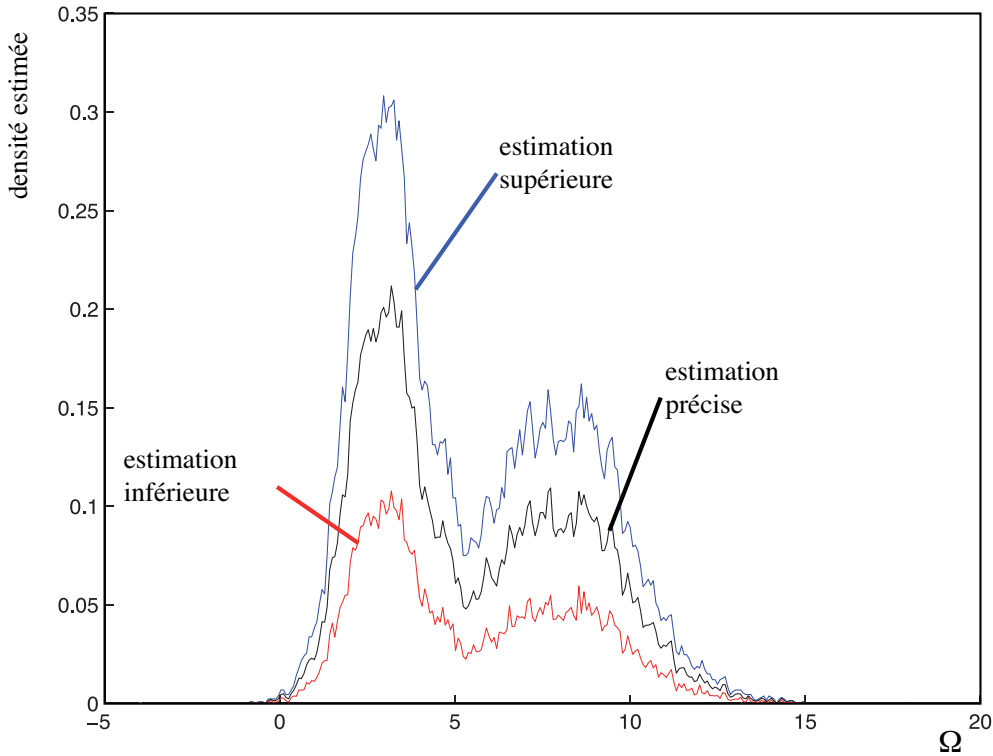
La seconde expérience a pour but est d'illustrer la Propriété III.19. La Figure IV.12 montre la superposition de l'estimation précise  $\hat{f}_{\kappa_\Delta}^n$  et l'estimation imprécise  $\bar{f}_{(\kappa_\Delta)}^n$  de la densité sous-jacente à cet ensemble d'observations en choisissant pour  $\kappa_\Delta$  le noyau d'Epanechnikov défini par l'expression (III.2.3). Le couple de noyaux maxitifs, dominant la décomposition sommative de  $\kappa_\Delta$ , le plus spécifique  $(\pi_\Delta^-, \pi_\Delta^+)$  est défini par les expressions (IV.3) et (IV.4). Sur la Figure IV.12, nous avons superposé l'estimation précise

$\hat{f}_{\kappa_{\Delta}}^n$  (en noir) ainsi que la borne inférieure (en rouge) et la borne supérieure (en bleu) de l'estimation imprécise  $\underline{f}_{(\kappa_{\Delta})}^n$ . On voit que  $\hat{f}_{\kappa_{\Delta}}^n$  est confondu avec  $\underline{f}_{(\kappa_{\Delta})}^n$ .



**Figure IV.12** : Superposition de l'estimation précise (noyau d'Epanechnikov) et l'estimation imprécise de la densité (domination la plus spécifique).

En gardant le noyau d'Epanechnikov défini par l'expression (I.32), l'examen de la Figure IV.12 montre que l'utilisation du couple de noyaux maxitifs (moins spécifique)  $(\pi_{\Delta}^-, \pi_{\Delta}^+)$  défini par les expressions (IV.5) et (IV.6), dominant la décomposition sommative de  $\kappa_{\Delta}$ , nous a amené à une estimation précise incluse dans l'estimation imprécise. On a, pour chaque valeur  $\{y_i\}_{i \in \{1, \dots, 500\}}$  de  $\Omega$ ,  $\hat{f}_{\kappa_{\Delta}}^n(y_i) \in \underline{f}_{(\kappa_{\Delta})}^n(y_i)$ .



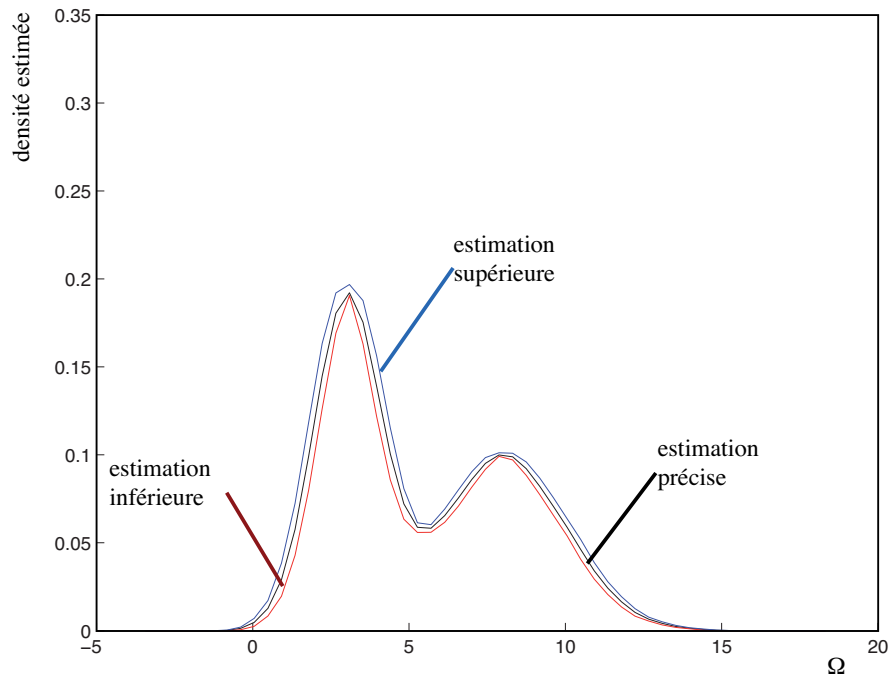
**Figure IV.13 :** Superposition de l'estimation précise (noyau d'Epanechnikov) et l'estimation imprécise de la densité (domination moins spécifiques).

## IV.2.2 Estimation précise et imprécise de la densité : cas particuliers

Nous proposons, dans cette partie, deux expériences illustrant le Théorème III.18 et Théorème III.29. Le Théorème III.18 dit que l'estimateur de densité de Parzen-Rosenblatt  $\hat{f}_{\kappa_\Delta}^n$  est inclus dans l'estimation imprécise  $\underline{f}_{[\kappa_\Delta]}^n$  défini par l'expression (III.60), obtenue en considérant que les deux noyaux de la décomposition sommative de la dérivée première de  $\kappa_\Delta$  peuvent être déduits l'un de l'autre par simple translation. De même, le Théorème III.29 dit que l'estimateur de Parzen-Rosenblatt  $\hat{f}_{\kappa_\Delta}^n$  est inclus dans l'estimation imprécise  $\underline{f}_{[\kappa_\Delta]^*}^n$  défini par l'expressions (III.76), obtenue en considérant que le noyau  $\kappa_\Delta$  est symétrique.

Pour ces deux expériences, nous avons simulé 1000 observations  $\{x_i\}_{i \in \{1, \dots, 1000\}}$ . L'intervalle de référence  $\Omega = [-5, 20]$  est subdivisé en 500 échantillons régulièrement espacés  $\{y_i\}_{i \in \{1, \dots, 500\}}$ . Le noyau utilisé est le noyau cosinus symétrisé défini par l'expression (IV.7) dont la largeur de bande  $\Delta = 1$ . Tandis que le noyau maxitif le plus spécifique, dominant la décomposition sommative de  $\kappa_\Delta$ , est défini par l'expression (IV.15).

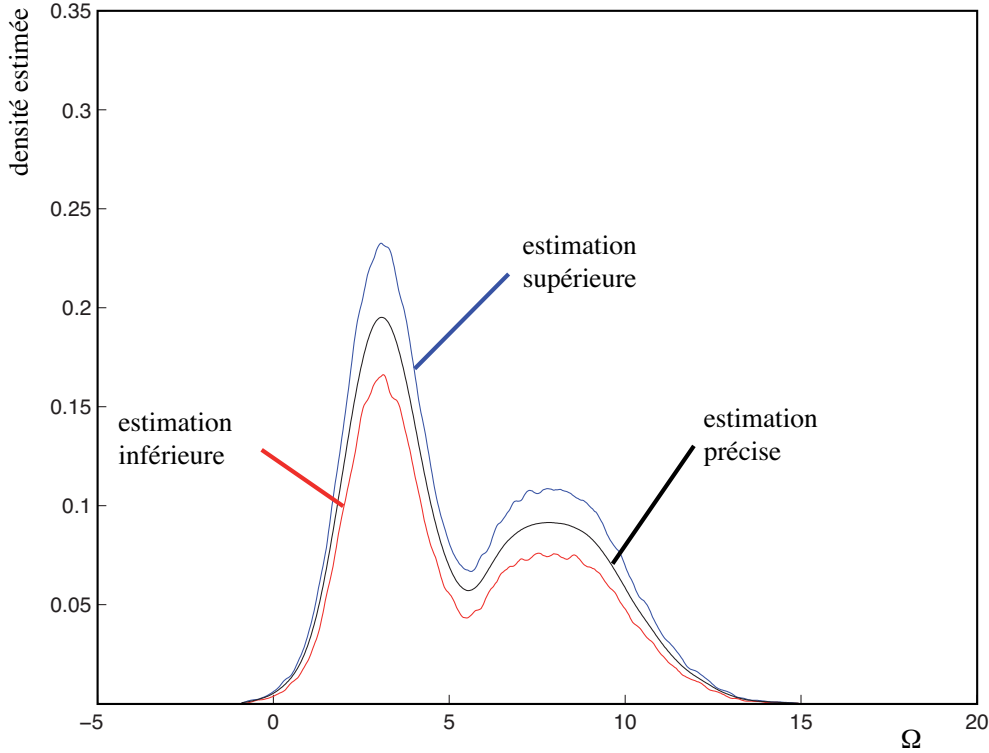
La première expérience illustre le Théorème III.18. La Figure IV.14 montre la superposition de l'estimation précise  $\hat{f}_{\kappa_\Delta}^n$  et l'estimation imprécise  $\underline{f}_{[\kappa_\Delta]}^n$ . Sur la Figure IV.14, l'estimation précise  $\hat{f}_{\kappa_\Delta}^n$  est dessinée en noir, la borne inférieure de l'estimation imprécise  $\underline{f}_{[\kappa_\Delta]}^n$  en rouge et sa borne supérieure en bleu. On peut voir que, pour chaque valeur  $\{y_i\}_{i \in \{1, \dots, 500\}}$  de  $\Omega$ ,  $\hat{f}_{\kappa_\Delta}^n(y_i) \in \underline{f}_{[\kappa_\Delta]}^n(y_i)$ .



**Figure IV.14 :** Première restriction : superposition de l'estimation précise (noyau cosinus symétrisé) et l'estimation imprécise (domination plus spécifique).

La seconde expérience illustre le Théorème III.29. Sur la Figure IV.14 nous avons superposé l'estimation précise  $\hat{f}_{\kappa_{\Delta}}^n$  (en noir), la borne inférieure de l'estimation imprécise  $\underline{\bar{f}}_{[\kappa_{\Delta}]^*}^n$  (en rouge) et sa borne supérieure (en bleu). La Figure IV.15 montre que, pour chaque valeur  $\{y_i\}_{i \in \{1, \dots, 500\}}$  de  $\Omega$ ,  $\hat{f}_{\kappa_{\Delta}}^n(y_i) \in \underline{\bar{f}}_{[\kappa_{\Delta}]^*}^n(y_i)$ .





**Figure IV.15 :** Deuxième restriction : superposition de l'estimation précise (noyau cosinus symétrisé) et l'estimation imprécise (domination plus spécifique).

### IV.2.3 Intégrale de l'imprécision de l'estimation imprécise de la densité

Nous proposons, dans cette partie, d'illustrer le fait que l'imprécision  $\Gamma_{(\kappa_\Delta)}^n$  définie par l'expression (III.36) de l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$  (III.31), ne dépend ni de la largeur de bande  $\Delta$  du noyau sommatif  $\kappa_\Delta$  ni du nombre d'observations  $n$ , comme le spécifie le Théorème III.20.

Dans cette expérience, nous avons choisi le noyau cosinus symétrisé défini par l'expression (IV.7) comme noyau de référence. La décomposition sommative de la dérivée première de  $\kappa_\Delta$  est dominée par la méthode objective. Les deux noyaux maxitifs  $\pi_\Delta^-$  et  $\pi_\Delta^+$ , intervenant dans cette domination, sont définis par les expressions (IV.10) et (IV.11). L'intégrale de  $\Gamma_{(\kappa_\Delta)}^n$  l'imprécision de l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$ , notée  $\varsigma(\kappa_\Delta)$ , est donnée par le Théorème III.20 :

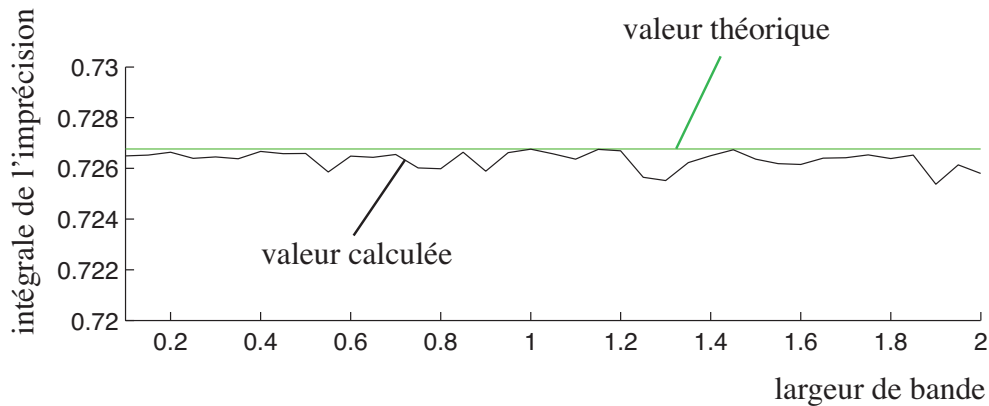
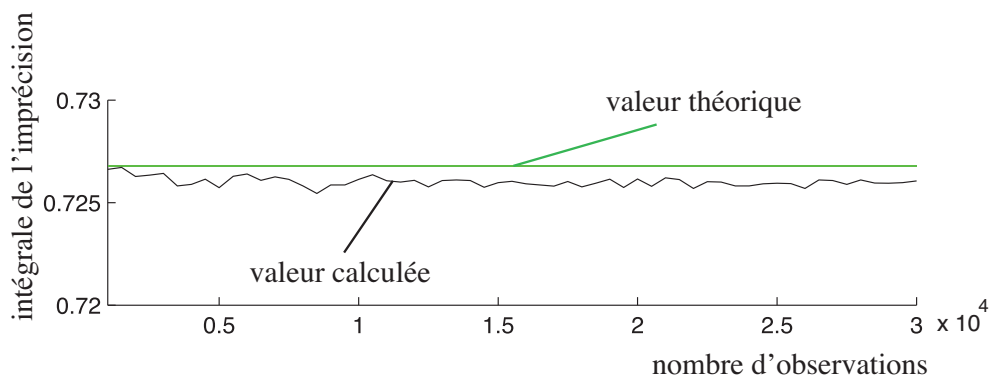
$$\begin{aligned}
 \varsigma(\kappa_\Delta) &= \int_{\Omega} \Gamma_{(\kappa_\Delta)}^n(u) du \\
 &= a(\rho(\pi_\Delta^-) + \rho(\pi_\Delta^+)), \text{ par la Définition (II.73)} \\
 &= a\left(\int_{-1}^0 \pi_\Delta^-(u) du + \int_0^1 \pi_\Delta^+(u) du\right) \\
 &\approx 0.7268,
 \end{aligned}$$

avec  $a = 1$  (voir Calcul (IV.4)) et

$$\begin{aligned}
\int_0^1 \pi_{\Delta}^+(u) du &= \int_{-1}^0 \pi_{\Delta}^-(u) du \\
&= \int_{-1}^{-0.5} (1 + \cos(\pi u)) du + \int_{-0.5}^0 (1 - \cos(\pi u)) du \\
&\approx 0.3634.
\end{aligned}$$

L'intégrale de l'imprécision de  $\bar{f}_{(\kappa_{\Delta})}^n$  est calculée d'une part en fixant  $n$  ( $n = 2000$ ) tandis que  $\Delta$  varie de 0.1 à 2 (Figure IV.16 (a)), d'autre part en fixant  $\Delta$  ( $\Delta = 1$ ) tandis que  $n$  varie de 1000 à 30000 (Figure IV.16 (b)).

La Figure IV.16 (a) (rsp. Figure IV.16 (b)) montre l'intégrale de l'imprécision en fonction de la largeur de bande (rsp. du nombre d'observations). Sur la Figure IV.16, la valeur théorique de l'intégrale de l'imprécision est dessinée en vert et sa valeur calculée en noir. On peut noter sur les deux Figures que la valeur de l'intégrale de l'imprécision est à peu près constante et fluctue autour de la valeur théorique.

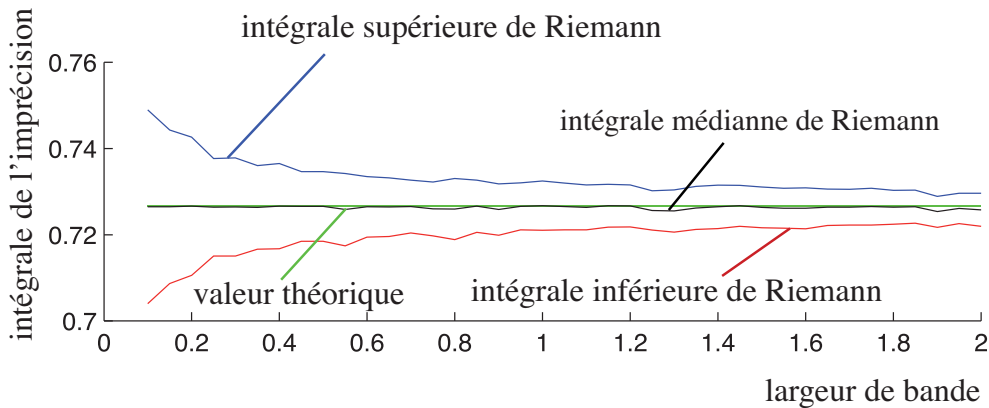
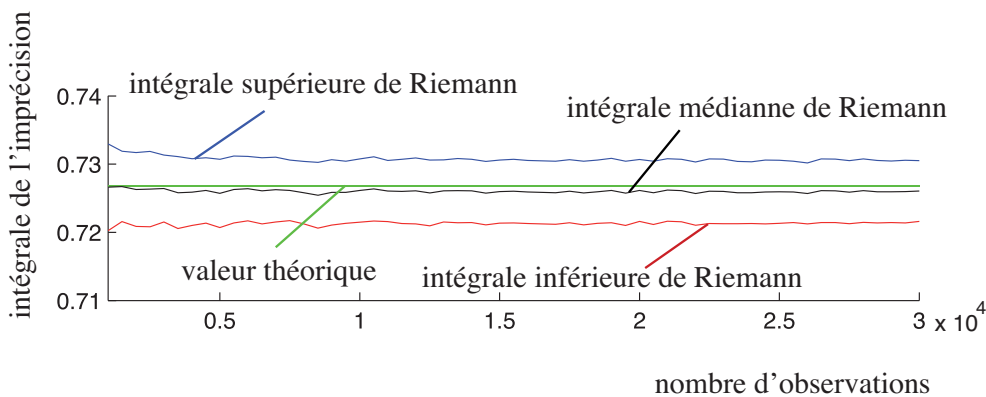
(a) variation de  $\Delta$ (b) variation de  $n$ 

**Figure IV.16** : Valeur théorique et valeur calculée de l'intégrale de l'imprécision en fonction : (a) de la largeur de bande et (b) du nombre d'observations.

Les fluctuations de la valeur de l'intégrale de l'imprécision sont dues aux erreurs numériques induites par l'utilisation de l'intégrale de Riemann discrète. Ce calcul se fait en utilisant les sommes de Riemann qui sont des approximations des intégrales.

Pour évaluer l'impact de ce calcul sur le résultat nous proposons de remplacer ce calcul précis mais incertain par un calcul imprécis mais certain. Pour obtenir la *borne inférieure de l'intégrale de Riemann*, on calcule  $\underline{S} = \sum_{i=1}^{499} \min(\Gamma_{(\kappa_\Delta)}^n(y_i), \Gamma_{(\kappa_\Delta)}^n(y_{i+1}))(y_i - y_{i+1})$ . La *borne supérieure de l'intégrale de Riemann*, notée  $\overline{S}$ , est obtenue par :  $\overline{S} = \sum_{i=1}^{499} \max(\Gamma_{(\kappa_\Delta)}^n(y_i), \Gamma_{(\kappa_\Delta)}^n(y_{i+1}))(y_i - y_{i+1})$ . La *médiane de l'intégrale de Riemann*, notée  $\tilde{S}$ , est donnée par :  $\tilde{S} = \frac{(\underline{S} + \overline{S})}{2}$ .

Sur la Figure IV.17, la valeur théorique de l'intégrale de l'imprécision est dessinée en vert, la somme médiane de Riemann en noir et la somme inférieure (rsp. supérieure) de Riemann en rouge (rsp. bleu). Les Figures IV.17 (a) et IV.17 (b) montrent la même expérience que celle illustrée par les Figures IV.16 (a) et IV.16 (b) en prenant en considération l'erreur d'approximation numérique. On peut voir clairement sur la Figure IV.17 que la valeur théorique de  $\zeta(\kappa_\Delta)$  appartient à l'intervalle formé par les approximations inférieures et supérieures de l'intégrale de Riemann.

(a) variation de  $\Delta$ (b) variation de  $n$ 

**Figure IV.17 :** Valeur théorique et valeur calculée de l'intégrale médiane, inférieure et supérieure de l'imprécision de l'estimation imprécise en fonction : (a) de la largeur de bande et (b) du nombre d'observations.

## IV.2.4 Caractérisation empirique de l'estimation imprécise de la densité

Nous proposons, dans cette partie, une expérience tendant à illustrer empiriquement deux indices de qualités de l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$  (III.31) de la densité sous-jacente à un ensemble fini d'observations. Le premier indice montre l'aptitude de l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$ , obtenue avec une seule expérience, à contenir la densité convoluée  $f_{\kappa_\Delta}$  définie par l'expression (III.41). Le second montre la corrélation entre l'imprécision  $\Gamma_{(\kappa_\Delta)}^n$  (III.36) de  $\underline{f}_{(\kappa_\Delta)}^n$  et la variance de l'estimateur de densité de Parzen-Rosenblatt  $\text{var}(\hat{f}_{\kappa_\Delta}^n)$ .

L'expérience se déroule de la manière suivante. On effectue 500000 tirages issus de la loi bimodale simulée. On divise l'ensemble de ces observations en 100 échantillons de 5000 observations. Pour chacun des 100 échantillons, on dispose, d'une part, d'une estimation précise  $\hat{f}_{\kappa_\Delta}^n$  en utilisant le noyau d'Epanechnikov défini par l'expression (III.2.3). Nous avons choisi ce noyau parce que c'est le noyau le plus utilisé en estimation non-paramétrique de densité. On dispose, d'autre part, d'une estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$  en utilisant le couple de noyaux maxitifs  $(\pi_\Delta^-, \pi_\Delta^+)$  défini par les expressions (IV.5) et (IV.6). Notons qu'une étude préalable du comportement de l'estimateur nous a permis de déterminer que la largeur de bande optimale, c'est-à-dire celle minimisant la distance MISE pour 5000 observations, est de 0.8. Ces estimations sont évaluées en 100 points d'échantillonnage régulièrement espacés  $\{y_i\}_{i \in \{1, \dots, 100\}}$  de  $\Omega = [-5, 20]$ . En chaque point d'échantillonnage  $\{y_i\}_{i \in \{1, \dots, 100\}}$ , on calcule, d'une part, la variance  $\text{var}(\hat{f}_{\kappa_\Delta}^n)(y_i)$  de l'estimation précise  $\hat{f}_{\kappa_\Delta}^n(y_i)$  et, d'autre part, pour chaque groupe d'observations, l'imprécision  $\Gamma_{(\kappa_\Delta)}^n(y_i)$  de l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n(y_i)$ .

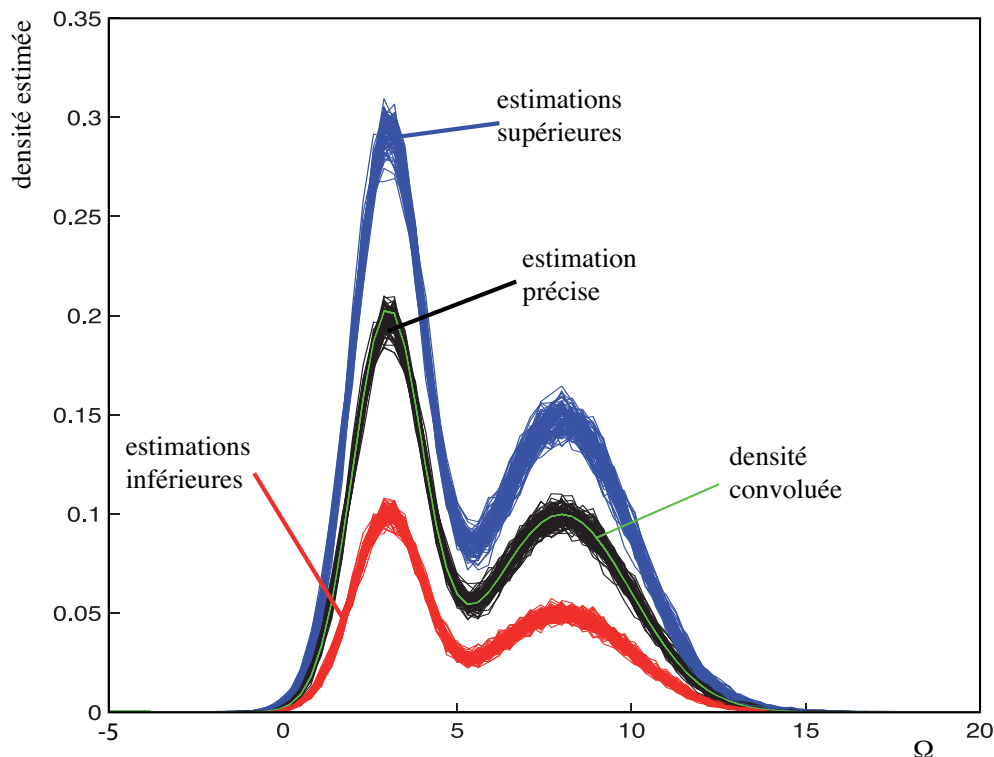
Nous calculons les indices de corrélations de Pearson<sup>2</sup>, Spearman<sup>3</sup> et Kendall<sup>4</sup> entre ces deux séries de valeurs pour des largeurs de bandes variant entre 0.2 et 1.6. La Figure IV.18 montre la superposition des estimations précises (en noir), la densité convoluée (en vert), la borne inférieure de l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$  (en rouge) et sa borne supérieure (en bleu). On peut voir sur la Figure IV.18 que, pour chaque valeur  $\{y_i\}_{i \in \{1, \dots, 100\}}$  de  $\Omega$ ,  $f_{\kappa_\Delta}(y_i) \in \underline{f}_{(\kappa_\Delta)}^n(y_i)$  quand le nombre d'observations  $n$  augmente.

---

2. Le coefficient de corrélation linéaire simple de deux variables  $X$  et  $Y$ , dit de Bravais-Pearson (ou de Pearson), noté  $r_p$ , est une normalisation de la covariance de deux variables par le produit des écarts-type de deux variables :  $r_p = \frac{\text{Cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$  ([130] pages 420-421).

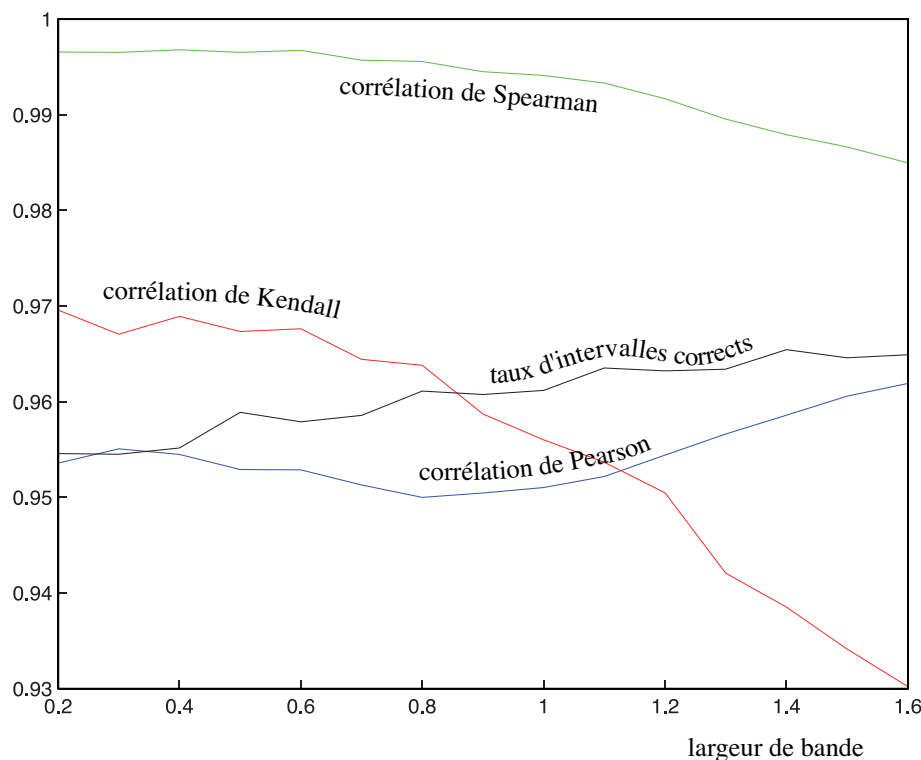
3. Le coefficient de corrélation de Spearman, noté  $r_s$ , se base sur le calcul du rang de chaque élément dans la série croissante de valeurs de variable  $X$  et de variable  $Y$  puis de calculer la différence de classement  $d_i$  où  $i \in \mathbb{N}$  dénoté le  $i$ -ème élément de l'échantillon. L'indicateur de Spearman est donné par :  $r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)}$  ([58] page 217).

4. Le tau de Kendall, noté  $\tau$ , mesure la force d'association entre deux variables aléatoires  $X$  et  $Y$ . Une valeur positive de  $\tau$  suggère la présence d'une dépendance positive entre les variables. Ce tau est défini par :  $\tau = \frac{N_c - N_d}{\frac{1}{2}n(n-1)}$  où  $N_c$  (resp.  $N_d$ ) est le nombre de paires concordantes (resp. discordante) dans un échantillon de taille  $n \in \mathbb{N}$ . Les paires d'observations  $(x_1, y_1)$  et  $(x_2, y_2)$  sont *concordantes* si et seulement si  $(x_1 > x_2 \text{ et } y_1 > y_2)$  ou  $(x_1 < x_2 \text{ et } y_1 < y_2)$ . Ces paires d'observations sont *discordantes* si et seulement si  $(x_1 > x_2 \text{ et } y_1 < y_2)$  ou  $(x_1 < x_2 \text{ et } y_1 > y_2)$  [59].



**Figure IV.18** : Superposition des 100 estimations précises (noyau d'Epanechnikov), estimations imprécises (domination moins spécifique) et densités convoluées.

Nous nous intéressons ensuite à l'aptitude de l'intervalle  $\overline{f}_{(\kappa_\Delta)}^n$  à contenir la vraie densité  $f_{\kappa_\Delta}$ . Pour caractériser cette aptitude, nous calculons, en chaque point  $\{y_i\}_{i \in \{1, \dots, 100\}}$ , le nombre d'intervalles, obtenu grâce à un échantillon, contenant la vraie valeur de  $f_{\kappa_\Delta}$ . Ce nombre est rapporté au nombre de valeurs testées pour obtenir le taux d'intervalle de prédictions correctes.



**Figure IV.19** : Courbes de performance de l'estimation imprécise en fonction de la largeur de bande.

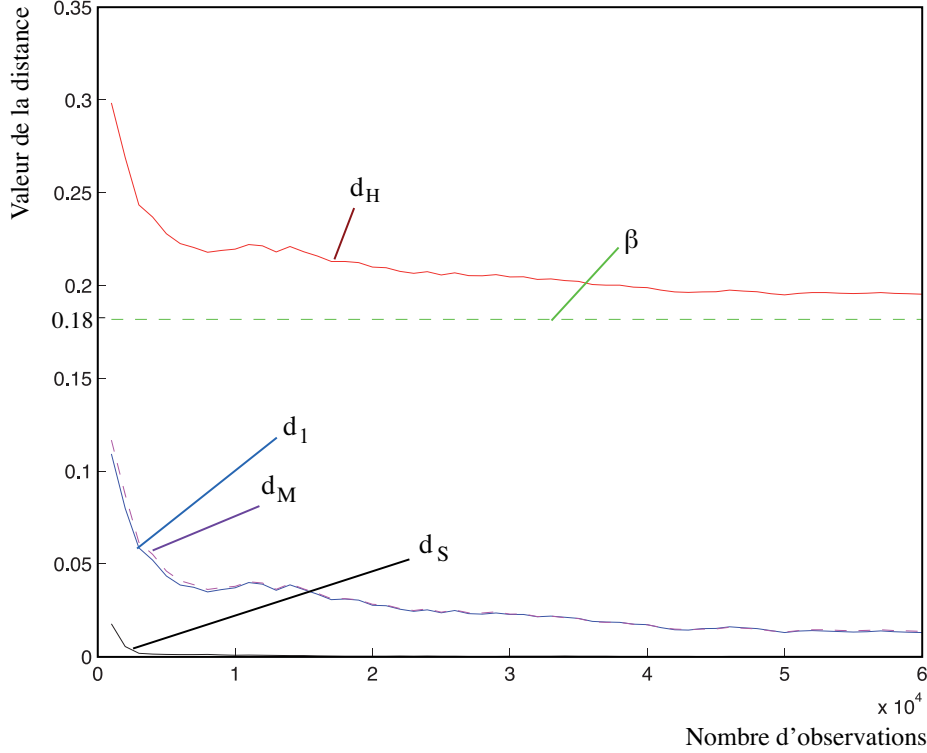
La Figure IV.19 montre l'évolution de ces quatre indices pour différentes largeurs de bande. Concernant les indices de corrélation, on peut constater une très forte corrélation entre la variance d'estimation précise de la densité et l'imprécision de l'estimation imprécise de la densité. Cette corrélation décroît lorsqu'on s'écarte de la largeur de bande optimale, mais reste cependant très élevée (supérieure à 0.8 dans la pratique). De même, on peut constater, via l'indice de cohérence, la très bonne aptitude de l'estimation intervalle à prédire la vraie densité, même lorsque la largeur de bande n'est pas optimale.

#### IV.2.5 Inclusions de la densité convoluée dans l'estimation imprécise de la densité : cas général

Nous proposons, dans cette partie, une expérience illustrant la convergence, en un certain sens, de l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$  de la densité vers la densité convoluée  $f_{\kappa_\Delta}$  (III.41) quand le nombre d'observations  $n$  tend vers l'infini, comme spécifie la Propriété III.22. Dans cette expérience, nous avons considéré le noyau cosinus symétrisé défini par l'expression (III.41) comme noyau de référence. Nous avons réalisé l'estimation imprécise  $\underline{f}_{(\kappa_\Delta)}^n$  en utilisant le couple de noyaux maxitifs le plus spécifique  $(\pi_\Delta^-, \pi_\Delta^+)$  (défini par les expressions (IV.10) et (IV.11)). Nous avons fixé la largeur de bande à  $\Delta = 1$  tandis que le nombre d'observations  $n$  varie de 1000 à 60000. L'intervalle de référence  $\Omega = [-5, 20]$  est subdivisé en 500 échantillons régulièrement espacés  $\{y_i\}_{i \in \{1, \dots, 500\}}$ .

Nous proposons d'illustrer cette convergence en utilisant les quatre distances  $d_H$ ,  $d_S$ ,  $d_1$  et  $d_M$  (cf. section III.2.4.3). La Figure IV.20 présente la superposition de la distance de Hausdorff  $d_H(f_{\kappa_\Delta}, \underline{f}_{(\kappa_\Delta)}^n)$  (III.48) (en rouge), la distance médiane  $d_M(f_{\kappa_\Delta}, \underline{f}_{(\kappa_\Delta)}^n)$  (III.57)

(en magenta pointillée), l'extension naturelle de la distance  $L^1$ ,  $d_S(f_{\kappa_\Delta}, \bar{f}_{(\kappa_\Delta)}^n)$  (III.53) (en noir) et la distance  $L^1$ ,  $d_1(f_{\kappa_\Delta}, \hat{f}_{\kappa_\Delta}^n)$  (III.44) (en bleu). Sur la même figure, la valeur minimale  $\beta$  (III.49) de  $d_H(f_{\kappa_\Delta}, \bar{f}_{(\kappa_\Delta)}^n)$  est dessinée en vert pointillé.



**Figure IV.20 :** Superposition de quatre distances  $d_H$ ,  $d_M$ ,  $d_S$  et  $d_1$  et la valeur constante  $\beta$  en fonction du nombre d'observations.

On peut voir sur la Figure IV.20, que  $d_H(f_{\kappa_\Delta}, \bar{f}_{(\kappa_\Delta)}^n)$  converge vers sa valeur minimale  $\beta = 0.18$  et  $d_S(f_{\kappa_\Delta}, \bar{f}_{(\kappa_\Delta)}^n)$  converge vers zero plus vite que les autres distances. La Figure IV.20 montre clairement que  $d_1(f_{\kappa_\Delta}, \hat{f}_{\kappa_\Delta}^n)$  et  $d_M(f_{\kappa_\Delta}, \hat{f}_{\kappa_\Delta}^n)$  sont presque complètement confondues, c'est-à-dire quand  $n$  tend vers l'infini,  $\hat{f}_{\kappa_\Delta}^n$  et  $\bar{f}_{\kappa_\Delta}^n$  convergent, de la même manière, vers la densité convoluée  $f_{\kappa_\Delta}$ . En ce sens, la fonction médiane  $\hat{f}_{\kappa_\Delta}^n$  peut être considérée comme un estimateur de densité de Parzen-Rosenblatt.

## IV.2.6 Lien entre la variance de l'estimation précise et l'imprécision de l'estimation imprécise

Nous proposons, dans cette partie, d'illustrer le fait que l'imprécision  $\Gamma_{(\kappa_\Delta)}^n$  (III.36) de l'estimation imprécise  $\bar{f}_{(\kappa_\Delta)}^n$  peut être utilisée pour estimer la variance  $var(\hat{f}_{\kappa_\Delta}^n)$  de l'estimation précise  $\hat{f}_{\kappa_\Delta}^n$ .

En théorie  $var(\hat{f}_{\kappa_\Delta}^n)$  décroît quand  $n$  et  $\Delta$  croient. Un résultat important issu de ([104] pages 38-40) donne :

$$\forall u \in \Omega, var(\hat{f}_{\kappa_\Delta}^n(u)) \approx (n\Delta)^{-1} f(u) R(\kappa_\Delta), \quad (\text{IV.16})$$

avec  $R(\kappa_\Delta) = \int_\Omega (\kappa_\Delta(u))^2 du$ .

Comme nous l'avons noté en section III.2.4.2, l'intégrale de l'imprécision  $\Gamma_{(\kappa_\Delta)}^n$  ne dépend ni de  $n$  ni de  $\Delta$ . La valeur de  $\Gamma_{(\kappa_\Delta)}^n$  ne peut donc pas être directement utilisée pour estimer  $\text{var}(\hat{f}_{\kappa_\Delta}^n)$ . Nous conjecturons qu'il suffit de multiplier cette imprécision par un facteur  $\gamma(n, \Delta)$  dépendant systématiquement de  $n$  et  $\Delta$ . La relation serait alors :

$$\forall u \in \Omega, \text{var}(\hat{f}_{\kappa_\Delta}^n(u)) = \gamma(n, \Delta) \mathbb{E}(\Gamma_{(\kappa_\Delta)}^n(u)). \quad (\text{IV.17})$$

En intégrant l'expression (IV.17), on a d'une part d'après l'expression (IV.16) :

$$\begin{aligned} \int_\Omega \text{var}(\hat{f}_{\kappa_\Delta}^n(u)) du &= \int_\Omega (n\Delta)^{-1} R(\kappa_\Delta) f(u) du \\ &= (n\Delta)^{-1} R(\kappa_\Delta), \end{aligned}$$

avec  $\int_\Omega f(u) du = 1$ ,

et d'autre part

$$\begin{aligned} \int_\Omega \gamma(n, \Delta) \mathbb{E}(\Gamma_{(\kappa_\Delta)}^n(u)) du &= \gamma(n, \Delta) \mathbb{E}\left(\int_\Omega \Gamma_{(\kappa_\Delta)}^n(u) du\right) \\ &= \gamma(n, \Delta) \varsigma(\kappa_\Delta), \end{aligned}$$

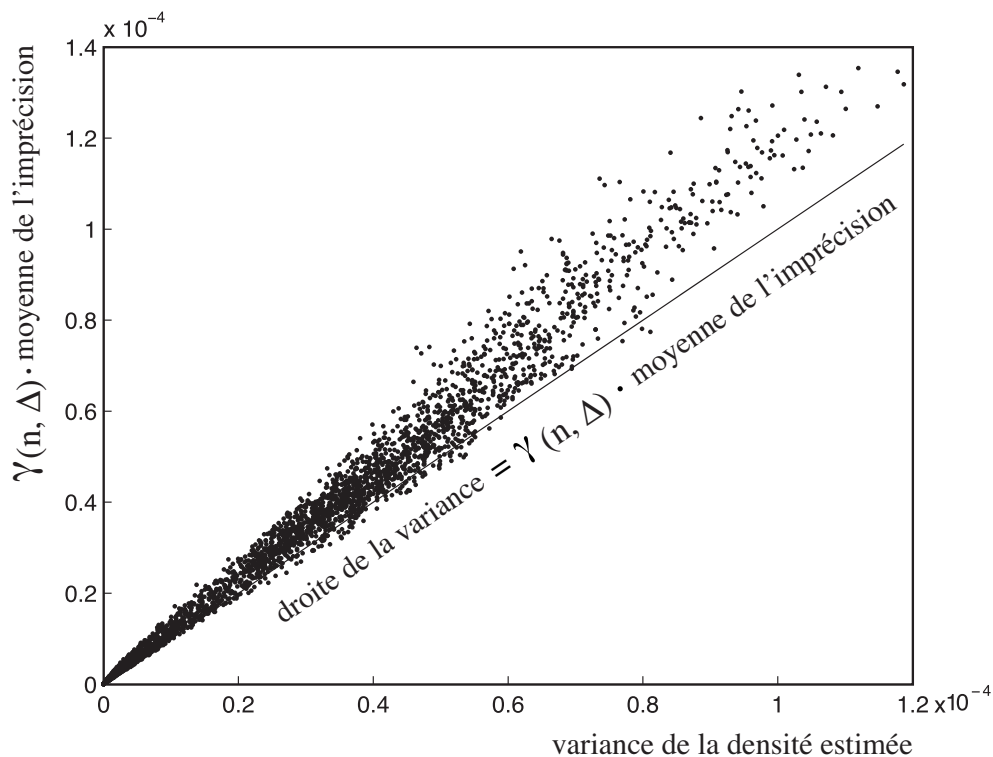
avec  $\int_\Omega \Gamma_{(\kappa_\Delta)}^n(u) du = \varsigma(\kappa_\Delta)$ .

D'où

$$\gamma(n, \Delta) = \frac{R(\kappa_\Delta)}{\varsigma(\kappa_\Delta) n \Delta}. \quad (\text{IV.18})$$

Cette expérience vise à illustrer cette conjecture en montrant la corrélation entre  $(\gamma(n, \Delta) \Gamma_{(\kappa_\Delta)}^n)$  et  $\text{var}(\hat{f}_{\kappa_\Delta}^n)$ . Nous avons utilisé le noyau cosinus symétrisé (IV.7) comme noyau de référence. Nous avons fixé la largeur de bande à  $\Delta = 1$ , tandis que le nombre d'observations  $n$  varie de 1000 à 10000. Pour chaque valeur de  $n$ , nous avons simulé 400 ensembles d'observations en utilisant toujours la même loi bimodale. Nous avons estimé la valeur de  $\text{var}(\hat{f}_{\kappa_\Delta}^n)$  et  $\gamma(n, \Delta) \mathbb{E}(\Gamma_{(\kappa_\Delta)}^n)$  sur 500 échantillons régulièrement espacés  $\{y_i\}_{i \in \{1, \dots, 500\}}$  de l'intervalle de référence  $\Omega = [-5, 20]$ . L'intégrale de l'imprécision est  $\varsigma(\kappa_\Delta) \approx 0.7268$  et la constante  $R(\kappa_\Delta) = \frac{3}{4\Delta} = 0.75$ .





**Figure IV.21** : Le nuage de valeurs de  $\gamma(n, \Delta)\mathbb{E}(\Gamma_{(\kappa\Delta)}^n)$  en fonction de  $\text{var}(\hat{f}_{\kappa\Delta}^n)$ .

La Figure IV.21 montre le résultat de cette expérience en dessinant  $\text{var}(\hat{f}_{\kappa\Delta}^n)$  en fonction  $\gamma(n, \Delta)\mathbb{E}(\Gamma_{(\kappa\Delta)}^n)$ . On peut voir sur la Figure IV.21 qu'il existe une forte corrélation entre  $\text{var}(\hat{f}_{\kappa\Delta}^n)$  et  $\gamma(n, \Delta)\mathbb{E}(\Gamma_{(\kappa\Delta)}^n)$  (coefficient de corrélation de Pearson  $r_p \approx 0.995$ ). Cependant, sur la Figure IV.21, le nuage de valeurs calculées est proche mais plutôt au dessus de la ligne correspondant à l'équation (IV.17). Ce décalage peut être expliqué soit par le fait que la relation (IV.16) est une approximation soit par le fait que la dépendance n'est peut être pas tout à fait linéaire.

# Conclusion générale

Les méthodes d'estimations non-paramétriques habituellement employées dans la littérature utilisent des techniques additives (distribution de probabilité) pour estimer une densité de probabilité sous-jacente à un ensemble fini d'observations. Le travail présenté dans ce mémoire propose de résoudre en partie ce problème en remplaçant l'estimation ponctuelle par une estimation intervalliste de probabilité. Cette estimation est en fait une enveloppe convexe d'un ensemble d'estimations de Parzen-Rosenblatt obtenues à l'aide d'une famille particulière de noyaux sommatifs. Cette méthode d'estimation imprécise de la densité est basée essentiellement sur une extension imprécise de la fonction de répartition développée par Loquin et Strauss [71]. Dans cette méthode, la borne inférieure (resp. supérieure) de l'intervalle d'estimation de la fonction de répartition est obtenue via l'intégrale de Choquet de la fonction de répartition empirique  $E_n$  par rapport à une mesure de nécessité (resp. possibilité) induite par un noyau maxitif [71]. Cette extension n'est pas directement applicable à l'estimation de la fonction de densité.

Nous proposons, dans ce mémoire, d'utiliser le lien entre la mesure empirique  $e_n$  et la fonction de répartition empirique  $E_n$  ( $e_n$  est la dérivée, au sens de distribution, de  $E_n$ ) pour bâtir directement un estimateur imprécis de la densité. Pour réaliser cette estimation, nous avons commencé par reformuler l'estimateur à noyau. Nous avons relié cette reformulation à la théorie de distribution de Schwartz [98], nous permettant ainsi de reformuler l'estimateur de densité comme une combinaison linéaire de deux estimations de la fonction de répartition. L'extension intervalliste de cet estimateur devient alors trivial en se basant sur les travaux de Loquin et Strauss [71]. Nous avons proposé trois extensions différentes permettant de prendre en considération des particularités des noyaux utilisés. Nous avons donné des algorithmes pratique de calcul exact de ces estimations.

Dans ce mémoire, nous présentons aussi un certain nombre de propriétés théoriques de cet opérateur d'estimation intervalliste de densité. Nous avons montré que, pour tout  $\kappa \in \mathcal{K}(\Omega)$ , l'intervalle d'estimation de la densité  $\overline{f}_{(\kappa)}^n$  contient tous les estimateurs de densité à noyau  $\widehat{f}_\eta^n$  tel que  $\eta \in \mathcal{D}(a, (\pi^+, \pi^-))$  où  $\mathcal{D}(a, (\pi^+, \pi^-))$  est l'ensemble des noyaux sommatifs ayant une décomposition sommative formée de  $a$  et  $(\pi^+, \pi^-)$ . Néanmoins, nous n'avons pas donné une interprétation physique à cet ensemble des noyaux sommatifs. Nous avons montré que l'intégrale de l'imprécision de l'intervalle d'estimation de la densité ne dépend ni de la largeur de bande du noyau maxitif utilisé ni du nombre d'observations. Il en résulte que cet intervalle d'estimation ne converge pas vers la vraie densité de probabilité. Nous avons cependant montré un certain type de convergence de  $\overline{f}_{(\kappa)}^n$  au sens où, quand le nombre d'observations  $n$  tends vers l'infini, la densité convoluée  $f_\kappa$  est incluse dans  $\overline{f}_{(\kappa)}^n$ . Enfin, nous avons illustré une conjecture intéressante reliant l'intégrale de l'imprécision de l'intervalle d'estimation de la densité et la variance de l'estimateur à noyau. En ce sens notre méthode pourrait permettre de quantifier l'erreur de l'estimation de la densité par

une méthode à noyau.

L'estimateur imprécis  $\overline{f}_{(\kappa)}^n$ , défini par sa propre expression (III.31), ne converge pas vers la vraie densité. Si on multiplie les bornes inférieures et supérieures de cet intervalle d'estimation par  $\frac{1}{n\Delta}$ , peut-on obtenir une convergence au sens où, quand la largeur de bande  $\Delta$  tend vers zéros et  $n\Delta$  tend vers l'infini quand le nombre d'observations  $n$  tend vers l'infini,  $\overline{f}_{(\kappa)}^n \rightarrow f$ ? Il est à noter que cette multiplication doit prendre en compte l'inclusion de l'estimateur à noyau  $\widehat{f}_\eta^n$  dans son extension  $\overline{f}_{(\kappa)}^n$ . Il nous semble aussi intéressant d'étudier les propriétés des intervalles d'estimations  $\overline{f}_{[\kappa]}^n$  et  $\overline{f}_{[\kappa]^*}^n$  (par exemple, l'imprécision, la convergence, ...). Comme nous l'avons souligné, dans ce mémoire, que si l'on considère le noyau d'Epanechnikov (comme noyau de base de l'estimateur à noyau), l'intervalle d'estimation  $\overline{f}_{(\kappa)}^n$  contient une seule fonction de densité c'est sa borne inférieure. Dans ce contexte, une question se pose : peut-on trouver le pourcentage des fonctions de densités dans chaque intervalle d'estimation? Une autre perspective à donner à ces travaux repose sur une comparaison de notre méthode d'estimation intervalliste avec les bandes de confiance proposées par Rosenblatt et Bickel [8]. Dans ce mémoire, on s'est limité à des estimateurs de densités unidimensionnels : un travail important reste à mener pour étendre cette méthode d'estimation imprécise en multidimensionnelles.





# Bibliographie

- [1] S. Abou-Jaoudé. Sur une condition nécessaire et suffisante de  $l_1$  convergence presque complète de l'estimateur de la partition fixe pour une densité. *C.R. Acad. Paris Sér. A-B*, 283(16 :AII) :A1107–A1110, 1976.
- [2] M. Arefi, R. Viertl, and S. M. Taheri. Fuzzy density estimation. *Metrika*, 72, 2010.
- [3] A. Azzalini. A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68 :326–328, 1981.
- [4] P. Barbé and M. Ledoux. *Probabilité*. EDP Sciences, France, 2007.
- [5] R. G. Bartle. *A Modern Theory of Integration*. American Mathematical Society, 2001.
- [6] C. Baudrit. *Représentation et propagation de connaissances imprécises et incertaines : Application à l'évaluation des risques liés aux sites et sols pollués*. PhD thesis, Université Paul Sabatier-Toulouse III, 2005.
- [7] A. Berlinet and G. Biau. Estimation de densité et prise de décision. In *Décision et Reconnaissance de Formes en Signal*, ed. Lengellé, R., pages 141–179. Hermès, Paris, 2002.
- [8] P. Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimators. *Annals of Statistics*, 3(6) :1071–1095, 1973.
- [9] J.-M. Bony. *Cours d'analyse : théorie des distributions et analyse de Fourier*. Ellipses Marketing, 2001.
- [10] D. Bosq and J. Lecoutre. *Théorie de l'estimation fonctionnelle*. Paris Economica, 1987.
- [11] T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. Aggregation operators : properties, classes and construction methods. In *Aggregation operators : new trends and applications*, pages 3–104, Heidelberg, Germany, 2002. Physica-Verlag GmbH.
- [12] S. D. Chatterji. *Cours d'analyse*. Presses Polytechniques et Universitaires Romandes (PPUR), 2004.
- [13] G. Choquet. Théorie des capacités. *Annales de l'Institut Fourier*, 5 :131–295, 1953.
- [14] F. Comby. *Estimation du mouvement apparent majoritaire dans une séquence d'images vidéo par accumulation de votes bimodaux sur un histogramme approché*. PhD thesis, LIRMM-Université Montpellier II-France, 2001.
- [15] R. Coppi, M. A. Gil, and H. A. L. Kiers. The fuzzy approach to statistical analysis. *Comput. Stat. Data Anal.*, 51(1) :1–14, 2006.
- [16] I. Couso and L. Sánchez. Higher order models for fuzzy random variables. *Fuzzy Sets Systems*, 159(3) :237–258, 2008.

- [17] V. Danilov and G. Koshevoy. Cores of cooperative games, superdifferentials of function, and the minkowski difference of sets. *Mathematical Analysis and Applications*, 247 :1–14, 2000.
- [18] L. de Campos, J. Huete, and S. Moral. Probability intervals : a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2 :167–196, 1994.
- [19] G. De Cooman. Possibility theory, part I : Measure-and integral-theoretic groundwork ; part II : Conditional possibility ; part III : Possibilistic independence. *Int. J. of General Systems*, 25 :291–371, 1997.
- [20] F. Demenge. *Espaces fonctionnels : utilisation dans la résolution des équations aux dérivées partielles*. EDP Sciences, 2007.
- [21] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38 :325–339, 1967.
- [22] D. Denneberg. *Non Additive Measure and Integral*. Kluwer Academic Publishers, Dordrecht, 1994.
- [23] T. Denoeux. Application du modèle des croyances transférables en reconnaissance de formes. *Traitement du Signal*, 14(5) :443–451, 1995.
- [24] S. Destercke. *Représentation et combinaison d’informations incertaines : nouveaux résultats avec applications aux études de sûreté nucléaires*. PhD thesis, Université Paul Sabatier-Toulouse III, 2008.
- [25] S. Destercke, D. Dubois, and E. Chojnacki. Une approximation possibiliste de la combinaison d’ensembles aléatoires consonants et indépendants. In *Rencontres Francophones sur la Logique Floue et ses Applications (LFA), Lens (France), 16/10/08-17/10/08*, pages 278–285. Cépaduès Editions, 2008.
- [26] L. Devroye. *A Course in Density Estimation*. Birkhauser, Boston, 1987.
- [27] L. Devroye and L. Györfi. *Nonparametric Density Estimation : The L1 View*. John Wiley and Sons, 1984.
- [28] L. Devroye and G. Lugosi. A universally acceptable smoothing factor for kernel density estimates. *Annals of Statistics*, 24(6) :2499–2512, 1996.
- [29] L. Devrye and T. Wagner. *Nonparametric discrimination and density estimation*. Technical Report no. 183, University of Texas, Austin, 1976.
- [30] M. Donsker. Justification and extension of doob’s heuristic approach to the kolmogorov-smirnov theorems. *The Annals of Mathematical Statistics*, 23 :277–281, 1952.
- [31] J. Doob. Heuristic approach to the kolmogorov-smirnov theorem. *The Annals of Mathematical Statistics*, 20 :393– 403, 1949.
- [32] D. Dubois. Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis*, 51(1) :47–69, Nov. 2006.
- [33] D. Dubois and H. Prade. *Possibility Theory An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.
- [34] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence (Canada)*, 4(4) :244–264, 1988.

- [35] D. Dubois and H. Prade. *Théorie des possibilités : Applications à la représentation des connaissances en informatique*. Masson, 1988.
- [36] D. Dubois and H. Prade. Logique floue. In *Série Arago. Observatoire français des techniques avancées*, 1994.
- [37] D. Dubois, H. Prade, L. Foulloy, and G. Mauris. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10 :273–297, 2004.
- [38] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In R. Lowen and M. Roubens, editors, *Fuzzy Logic. State of the Art*, pages 103–112. Kuwer Acad. Publ., Dordrecht, 1993.
- [39] D. Dubois, H. Prade, and P. Smets. A definition of subjective possibility. *International Journal of Approximate Reasoning*, 48 :352–364, 2008.
- [40] V. A. Epanechnikov. Nonparametric estimation of a multidimensional probability density. *Theoretical probability, applications*, 14 :153–158, 1969.
- [41] M. Falk. Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statist. Neerlandica*, 37 :73–83, 1983.
- [42] S. Ferson, V. Kreinovich, L. Ginzburg, D. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures, sand2002-4015. Technical report, Sandia National Laboratories, Albuquerque, New Mexico, 2002.
- [43] E. Fix and J. Hodges. *Discriminatory analysis, nonparametric discrimination : consistency properties*. Report No. 4, Project No. 21-49-004, USAF School of Aviation Medicine, 1951.
- [44] J. Geoffrey. Sur l'estimation d'une densité dans un espace métrique. *C.R. Acad. Paris Sér. A*, 278 :1449–1452, 1974.
- [45] M. Gil. Fuzzy random variables : Development and state of the art. In *Mathematics of Fuzzy Systems, Linz Seminar on Fuzzy Set Theory*, pages 11–15. Linz, Austria, 2004.
- [46] M. A. Gil, M. López-Díaz, and D. A. Ralescu. Overview on the development of fuzzy random variables. *Fuzzy Sets and Systems*, 157(19) :2546–2557, 2006.
- [47] L. Ginzburg, R. Akçakaya, and S. Ferson. Whereof one cannot speak : When input distributions are unknown, 1996.
- [48] N. Glick. Consistency conditions for probability estimators and integrals of density estimators. *Utilitas Math.*, 23 :61–74, 1974.
- [49] G. González-Rodríguez, A. Colubi, and M. Ángeles Gil. A fuzzy representation of random variables : An operational tool in exploratory analysis and hypothesis testing. *Comput. Stat. Data Anal.*, 51(1) :163–176, 2006.
- [50] U. Grenander. On the theory of mortality measurement part II. *Skand. Akt.*, 39 :125–153, 1956.
- [51] P. Hall and J. S. Marron. Estimation of integrated squared density derivatives. *Statistics & Probability Letters*, 6(2) :109–115, 1987.
- [52] W. Härdle. *Smoothing Techniques, With Implementations in S*. Springer, 1991.
- [53] J. Istas. *Introduction aux modélisations mathématiques pour les sciences du vivant*. Springer, 2000.



- [54] V. Jeyakumar and A. M. Rubinov. *Continuous Optimization : Current Trends and Modern Applications (Applied Optimization)*. Springer ; 1 edition, 2005.
- [55] P. Johnson. *A History of Set Theory*. Prindle, Weber & Schmidt, 1972.
- [56] M. C. Jones. The performance of kernel density functions in kernel distribution function estimation. *Statistics & Probability Letters*, 37(2) :129–132, 1990.
- [57] M. C. Jones and S. J. Sheather. Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics & Probability Letters*, 11(6) :511–514, 1991.
- [58] M. J.S. *Distribution-Free Statistical Methods*. Chapman and Hall, New York, 1981.
- [59] M. Kendall. A new measure of rank correlation. *Biometrika*, 30 :81–89, 1938.
- [60] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic : Theory and Applications*. Prentice Hall, Englewood Cliffs, N.J, 1995.
- [61] A. Kolmogorov. Sulla determinazione empirica di una legge de distribuzione. *Giornale dell’Istituto Italiano degli Attuari*, 4 :83–91, 1933.
- [62] A. N. Kolmogorov. *Elements of the Theory of Functions and Functional Analysis*. Dover Publications, Boston, 1999.
- [63] R. Kruse and K. Meyer. *Statistics with Vague Data*. Plenum Press, 1987.
- [64] H. Kwakernaak. Fuzzy random variables. part I : Definitions and theorems. *Information Sciences*, 15(1) :1–29, 1978.
- [65] H. Kwakernaak. Fuzzy random variables. part II : algorithms and examples for the discrete case. *Information Sciences*, 17 :253–278, 1979.
- [66] U. Lall and Y. Moon. Kernel flood frequency estimators : Bandwidth selection and kernel choice. *Water Resour. Res.*, 29(4) :1003–1015, 1993.
- [67] J. Lecoutre. *Contribution à l’estimation non paramétrique de la régression*. PhD thesis, Université de Pierre et Marie Curie-ParisVI-France, 1982.
- [68] M. Lejeune. *Statistique : La théorie et ses applications*. Springer, France, 2004.
- [69] K. Loquin. *De l’utilisation des noyaux maxitifs en traitement de l’information*. PhD thesis, LIRMM-Université Montpellier II-France, 2008.
- [70] K. Loquin and O. Strauss. Fuzzy histograms and density estimation. In J. Lawry, E. Miranda, A. Bugarín, S. Li, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Soft Methods in Probability and Statistics*, pages 45–52. Springer, 2006.
- [71] K. Loquin and O. Strauss. Imprecise functional estimation : the cumulative distribution case. In *SMPS Soft Methods in Probability and Statistics*, pages 175–182, Toulouse, France, Sept. 2008.
- [72] K. Loquin and O. Strauss. On the granularity of summative kernels. *Fuzzy sets and systems*, 159 :1952–1972, 2008.
- [73] K. Loquin and O. Strauss. Noise quantization via possibilistic filtering. In *International Symposium on Imprecise Probability : Theory and Applications, ISIPTA 2009*, pages 297–306, Durham, UK, jul 2009.
- [74] M. K. Lothar Gaul and M. Wagner. *Boundary Element Methods for Engineers and Scientists : An Introductory Course With Advanced Topics*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2003.

- [75] G.-P. Mafouta-Bantsimba. *Mathématiques Pour L' économie*. De Boeck, 2005.
- [76] M.-H. Masson. *Apports de la théorie des possibilités et des fonctions de croyance à l'analyse de données imprécises*. Mémoire présenté en vue de l'obtention du diplôme d'Habilitation à Diriger des Recherches. Université de Technologie de Compiègne-France, 2 décembre 2005.
- [77] G. Matheron. *Random sets and integral geometry*. John Wiley and sons, New York, 1975.
- [78] I. Molchanov. *Theory of Random Sets*. Springer, 2005.
- [79] A. P. Morse. *A Theory of Sets*. Academic Press, Orlando, Florida, 1965.
- [80] T. Murofushi and M. Sugeno. Fuzzy measures and fuzzy integrals. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals : Theory and Applications*, pages 3–41. Physica Verlag, Heidelberg, 2000.
- [81] E. Nadaraya. Some new estimates for distribution function. *Theory of Probability and its Application*, 9 :497–500, 1964.
- [82] S. B. Nadler. *Hyperspaces of Sets*. Marcel Dekker, New York, 1978.
- [83] S. P. Niculescu and R. Viertl. Bernoulli's law of large numbers for vague data. *Fuzzy Sets Syst.*, 50(2) :167–173, 1992.
- [84] A. Novotny and I. Straskraba. *Introduction to the Mathematical Theory of Compressible Flow*. Oxford University Press, 2005.
- [85] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33 :1065–1076, 1962.
- [86] S. Petit-Renaud. *Application de la théorie des croyances et des systèmes flous à l'estimation fonctionnelle en présence d'informations incertaines ou imprécises*. PhD thesis, Université de Technologie de Compiègne. Laboratoire HeuDiaSyc-UMR-France, 1 décembre 1999.
- [87] M. Puri and D. Ralescu. Fuzzy random variables. *Journal of Mathematical Analysis and Applications*, 114 :409–422, 1986.
- [88] B. P. Rao. *Nonparametric Functional Estimation*. Academic Pr, 1983.
- [89] E. Raufaste. *Psychologie Ergonomique de l'optimisation des Processus Décisionnels*. Mémoire présenté en vue de l'obtention du diplôme d'Habilitation à Diriger des Recherches. Université Toulouse le Mirail U.F.R de Psychologie-France, 10 décembre 2003.
- [90] R. D. Reiss. Nonparametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics*, 8 :116–119, 1981.
- [91] M. B. Retali. Convergence uniforme d'un estimateur de la densité par la méthode de noyau. *Revue Roumaine Math. Pure et Appliquées*, 23 :361–385, 1978.
- [92] A. Rico and O. Strauss. Imprecise expectations for imprecise linear filtering. *International Journal of Approximate Reasoning*, 51 :933–947, 2010.
- [93] V. Rivoirard and G. Stoltz. *Statistique en action*. Vuibert, 2009.
- [94] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27 :832–837, 1956.

- [95] T. A. Runkler. Fuzzy histograms and fuzzy chi-squared tests for independence. In *IEEE international conference on fuzzy systems*, pages 1361–1366, 2004.
- [96] D. Schmeidler. Integral representation without additivity. In *Proceedings of the American Mathematical Society*, volume 97 of 2, pages 255–261, June 1986.
- [97] R. Schneider. *Convex Bodies : The Brunn-Minkowski Theory*. Cambridge University Press, Cambridge, 1993.
- [98] L. Schwartz. *Théorie des distributions*. Hermann, Paris, 1950.
- [99] D. Scott. Averaged shifted histograms : Effective nonparametric density estimation in several dimensions. *Annals of Statistics*, 13 :1024–1040, 1985.
- [100] D. W. Scott. *Multivariate Density Estimation : Theory, Practice, and Visualization*. Wiley Interscience, New York, 1992.
- [101] R. Servien. *Estimation de régularité locale*. PhD thesis, Université Montpellier II-France, 2010.
- [102] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [103] A. F. Shapiro. Fuzzy random variables. *Insurance : Mathematics and Economics*, 44(2) :307–314, 2009.
- [104] B. Silvermann. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.
- [105] C. Simon and P. Weber. Analyse de la fiabilité imprécise des systèmes par les réseaux de fonctions de croyance. In *JFRB'08 4èmes journées Francophones sur les Réseaux Bayésiens*, pages 105–119, Lyon, France, 2008.
- [106] J. S. Simonoff. *Smoothing Methods in Statistics*. Springer-Verlag, 1996.
- [107] R. S. Singh, T. Gasser, and B. Prasad. Nonparametric estimates of distributions functions. *Communication in Statistics - Theory and Methods*, 12 :2095–2108, 1983.
- [108] N. Smirnov. Approximate laws of distribution of random variables from empirical data. *Uspekhi Matematicheskikh Nauk*, 10 :179–206, 1944.
- [109] O. Strauss. Nouvelles représentations des mesures de confiances. In *7ème Journées Nationales de la Recherche en Robotique (JNRR 2009)*, pages 605–610, Neuvy-sur-Barangeon, France, novembre 2009.
- [110] O. Strauss, F. Comby, and M.-J. Aldon. Rough histograms for robust statistics. In *ICPR*, pages 2684–2687, 2000.
- [111] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, Japan., 1974.
- [112] J. Swanepoela. Mean integrated squared error properties and optimal kernels when estimating a distribution function. *Communications in Statistics-Theory and Methods*, 17(11) :3785–3799, 1988.
- [113] W. Trutschnig. Fuzzy probability distributions induced by fuzzy random vectors. In *SMPS*, pages 71–78, Bristol - UK, September 2006.
- [114] W. Trutschnig. A strong consistency result for fuzzy relative frequencies interpreted as estimator for the fuzzy-valued probability. *Fuzzy Sets Syst.*, 159(3) :259–269, 2008.
- [115] A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer science and business media, New York, 2009.

- [116] J. van den berg, U. Kaymak, and W.-M. van den bergh. Financial markets analysis by using a probabilistic fuzzy modelling approach. *Int. J. Approx. Reasoning*, 35(3) :291–305, 2004.
- [117] R. Viertl. Univariate statistical inference with fuzzy data. *Comput. Statist. Data Anal.*, 51 :133–147, 2006.
- [118] R. Viertl. On reliability estimation based on fuzzy lifetime data. *Journal of Statistical Planning and Inference*, 139 :1750–1755, 2009.
- [119] R. Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton, FL 1996.
- [120] R. Viertl and W. Trutschnig. Fuzzy histograms and fuzzy probability distributions. In *11th Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems*. EDK, 2006.
- [121] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.
- [122] L. Waltman, U. Kaymak, and J. Van Den Berg. Fuzzy histograms : A statistical analysis. In *EUSFLAT-LFA 2005 Joint 4th EUSFLAT 11th LFA Conference*, pages 605–610, Barcelona, september 2005.
- [123] G. Watson and M. Leadbetter. Hazard analysis II. *Sankhya Ser. A*, 26 :101–116, 1964.
- [124] B. Winter. Convergence rate of perturbed empirical distribution functions. *Journal of Applied Probability*, 16 :163–173, 1979.
- [125] R. R. Yager. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9 :249–260, 1983.
- [126] H. Yamato. Uniform convergence of an estimator of a distribution function. *Bulletin on Mathematical Statistics*, 15 :69–78, 1973.
- [127] L. Zadeh. Fuzzy sets. *Information and Control*, 8 :338–353, 1965.
- [128] L. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Inform.Sci.*, 8 :199–249, 1975.
- [129] L. Zadeh. Fuzzy sets as a basis for theory of possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.
- [130] H. Zuse. *A Framework of Software Measurement*. Walter de Gruyter, 1997.



## Résumé

Dans cette thèse, nous proposons une nouvelle méthode d'estimation non-paramétrique de la densité de probabilité. Cette méthode d'estimation imprécise combine la théorie de distribution de Schwartz et la théorie de possibilité. La méthode d'estimation que nous proposons est une extension de la méthode d'estimation à noyau. Cette extension est basée sur une nouvelle méthode de représentation de la notion de voisinage sur laquelle s'appuie l'estimation à noyau. Cette représentation porte le nom de noyau maxitif. L'estimation produite est de nature intervalliste. Elle est une enveloppe convexe d'un ensemble d'estimation de Parzen-Rosenblatt obtenus avec un ensemble de noyaux contenus dans une famille particulière.

Nous étudions un certain nombre des propriétés théoriques liées à cette nouvelle méthode d'estimation. Parmi ces propriétés, nous montrons un certain type de convergence de cet estimateur. Nous montrons aussi une aptitude particulière de ce type d'estimation à quantifier l'erreur d'estimation liée à l'aspect aléatoire de la distribution des observations. Nous proposons un certain nombre d'algorithmes de faible complexité permettant de programmer facilement les méthodes que nous proposons.

**Mots-clés :** Estimateur de densité de Parzen-Rosenblatt, noyaux sommatifs, espérance sommatif, théorie des distributions, théorie des possibilités, probabilités imprécises, noyaux maxitifs, espérance maxitive, intégrale de Choquet.

## Abstract

This manuscript, proposes a new nonparametric method for estimating the probability density function. This estimation method combines the Schwartz distribution theory and the possibility theory. It is an extension of the kernel density estimator that leads to imprecise estimation. It is based on a new method for modeling neighborhood. The interval valued estimate it produces is a convex envelope of the Parzen-Rosenblatt estimates obtained with kernels belonging to a coherent convex family.

We prove some theoretical properties of this new method. Among these properties, we have shown a kind of convergence of this estimator. We also shown a particular aptitude of this estimator to quantify the error due to random variation in observation. We also propose very low complexity algorithms to compute the proposed methods.

**Key-words :** Parzen-Rosenblatt estimator, summative kernels, summative expectation, distributions theory, possibility theory, imprecise probabilities, maxitive kernels, maxitive expectation, Choquet integral.