De l'utilisation des noyaux maxitifs en traitement de l'information

Kevin Loquin

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

Kevin Loquin

le 03 Novembre 2008

De l'utilisation des noyaux maxitifs en traitement de l'information

Président :	Mme. Valérie BERTHÉ *Directeur de recherche au LIRMM (Montpellier)*
Rapporteurs :	Mme. Isabelle BLOCH *Professeur à l'ENST (Paris)* M. Didier DUBOIS *Directeur de recherche à l'IRIT (Toulouse)*
Examinateurs :	M. Gérard BIAU *Professeur à l'UPMC (Paris)* M. Olivier STRAUSS *Maître de conférence au LIRMM (Montpellier)*

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
161 rue Ada - 34392 Montpellier Cedex 5 - France

Remerciements

Les travaux présentés dans ce mémoire de thèse ont été effectués au sein de l'équipe ICAR (Image et Interactions) au LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier).

Mes premiers remerciements vont, très chaleureusement, à Olivier Strauss, mon directeur de thèse. Olivier a su, pendant ces trois années, être l'encadrant idéal : il m'a aiguillé, orienté judicieusement quand c'était nécessaire, m'a laissé du mou et de la liberté quand il le fallait. Ses grandes compétences à la croisée de nombreux domaines applicatifs du traitement de l'information et des théories de l'incertain ainsi que sa rigueur scientifique se sont parfaitement alliées à mes compétences et mes attentes. Cette complémentarité a permis de donner les bonnes directions à nos recherches. Ce document doit beaucoup à Olivier Strauss et je dois également beaucoup à Olivier, qui m'a conforté dans mon souhait d'orienter ma vie professionnelle vers la recherche. Si je me sens épanoui dans ce métier, c'est pour beaucoup grâce à lui.

Isabelle Bloch, Professeur à l'ENST à Paris, et Didier Dubois, Directeur de recherche à l'IRIT à Toulouse, ont accepté d'être les rapporteurs de cette thèse, je les remercie vivement. J'ai apprécié leurs rapports et leurs remarques fructueuses sur ce document. Merci également à Valérie Berthé, Directeur de recherche au LIRMM à Montpellier, présidente de mon jury de thèse et à Gérard Biau, Professeur à l'UPMC à Paris, examinateur de ma thèse, dont les remarques m'ont permis de développer rigoureusement la partie "statistiques" de ce document.

Gérard Biau a fait partie de mon comité de thèse avec Bruno Tisseyre. Je les remercie pour leurs remarques et leurs aides pendant et hors des réunion de mon comité de thèse qui m'ont grandement apporté.

Du 2 au 8 Juillet 2008, j'ai organisé l'école d'été SIPTA sur les probabilités imprécises. J'ai collaboré, lors de l'organisation de cet évènement, avec Caroline Imbert, Céline Berger et Elisabeth Greverie. Ces trois personnes ont fait des merveilles pour cette école d'été, je ne les remercierai jamais assez. Je pense à bien d'autres personnes qui ont participé à l'organisation de cette école, notamment Jean-Marc Bernard, Chargé de recherche au LPE à Paris, Sébastien Destercke (mon alter ego toulousain, nouvellement chercheur au CIRAD), Étienne Dombre, Directeur de Recherche au LIRMM, que je tiens à remercier. Je souhaite également remercier la société SIPTA et Marco Zaffalon son président pour m'avoir fait confiance dans l'organisation de ce projet.

Je tiens également à remercier Marie-José Aldon et William Puech (pour qui le 3 Novembre est aussi une date importante), Maître de conférences, responsable du groupe ICAR, qui furent mes deux premiers directeurs de thèse, avant qu'Olivier Strauss ne soit habilité à diriger des recherches (HDR).

Je souhaite également remercier Florence Jacquey et Frédéric Comby pour leur collaboration scientifique. Je voudrais également remercier fortement Bilal Nehme pour nos collaborations et lui souhaiter bon courage pour ses futurs recherches en traitement de l'information avec les nouvelles théories de l'incertain.

Si je pense à ces trois années au LIRMM, naturellement me viennent des noms : Mourad Benoussaad (Moumou), Carla Aguiar (Caca), Milan Djilas (Spy), Ashvin Sobhee (Maurice), Ciao Liu (Richard), Rogerio Richa, Yousra Ben Zaida (Youyou), Lotfi Chikh (Lofti), Baptiste Magnier (Batou), Dingguo Zhang, Floor Campfens, Walid Zarrad, Peter Meuel (La meule), Arturo Gil (Pinto), Sébastien Lengagne, Jérémy Laforet, Philippe Amat, Aurélien Noce, Andreea-Elena Acunta, José Rodriguès, Gael Pages, Antonio Bo, Vincent Bonnet, Nicolas Riehl, Sébastien Cotton, David Corbel, Michel Dominici, Jean-Mathias Spiewak, Abdellah El Jalaoui, Maria Papaiordanidou, Olivier Parodi, Sébastien Durand (Dudu), Lama Mourad, Khizar Hayat, Zafar Shahid, Hai Yang, Lei Zhang, Mohamed Diaa Ahmadou, Nabil Zemiti. J'ai passé trois excellentes années au département robotique du LIRMM avec ces personnes (et avec ceux que j'oublie).

Je voudrais maintenant remercier les gens que j'aime : mes parents, mes grands parents et toute ma famille. Ils m'ont tous, chacun à leur manière, permis de devenir le docteur heureux que je suis. Merci vraiment du fond du cœur. Parmi les gens que j'aime, Alice tient une place toute particulière. Une place parfois ingrate auprès de moi mais place qui m'a permis, me permet et nous permet d'avancer. Merci.

Je souhaite également remercier Béatrice, Guy, Guillaume et Alexia présents à ma soutenance et maintenant dans ma vie.

Je remercie également Julien, mon soutien marseillais ainsi que tous mes amis de Rouen : Polo, Gaward, Timo, Ducduc, Élise et les autres. Merci également à Chanchan des champs et à Seb et Simon.

Merci à Bowie pour tes griffures...

Table des matières

Liste des tableaux

Table des figures

Introduction générale

Étymologiquement, le mot statistique signifie étude de données. Les statistiques mathématiques, telles qu'on les comprend aujourd'hui, s'appuient sur les probabilités pour déterminer les caractéristiques d'un ensemble de données. C'est à Adolphe Quételet, docteur en sciences mathématiques à l'université de Gand en 1819, que l'on doit ce lien entre études des données et probabilités. Avant lui, les statistiques étaient assimilées à de simples relevés de données comme, par exemple, des relevés des cours du bétail ou des recensements.

Le traitement du signal est beaucoup plus récent que les statistiques. Il s'appuie cependant sur le même terreau : les données. En statistiques on parle généralement de données relevées ou d'observations alors qu'en traitement du signal, on parle plutôt de données acquises ou de mesures. Le développement de l'informatique a été, au cours de ces dernières années, un accélérateur au développement du traitement du signal numérique. Voici quelques exemples d'applications du traitement numérique du signal :

- Télécommunications : téléphonie, transfert de données numériques par voie terrestre ou par satellite. Internet ou le GPS sont des exemples d'outils quotidiens qui doivent leur développement et leur prolifération aux avancées technologiques du traitement numérique du signal.
- Audio : amélioration des techniques d'enregistrement et de compression. Le traitement du son a changé de nature avec les ordinateurs : le CD ou les formats de compression mp3, wma ou autres sont numériques (alors que le vinyl était analogique).
- Imagerie : analyses à visées médicales (reconstruction tomographique, imagerie par résonance magnétique - IRM), algorithmes de traitement d'image pour des outils de création artistique tels que Photoshop©. Ce domaine inclut aussi les techniques de reconnaissance de formes et de compression.

Tout comme pour les statistiques, derrière toutes ces applications modernes du traitement du signal se cache la théorie des probabilités. Est-il possible de remplacer la sacrosainte théorie des probabilités ?

La théorie des possibilités est un cas particulier de la théorie des probabilités imprécises. Cette théorie mesure les incertitudes, non plus par des valeurs ponctuelles, mais par des valeurs intervallistes. La quantification de l'incertitude est ainsi rendu plus fiable, plus honnête. En effet, les données que l'on possède pour évaluer les incertitudes sont généralement imparfaites ; c'est-à-dire qu'elles peuvent être rares, vagues ou conflictuelles entre elles. Si c'est à Gand que les statistiques ont été reliées aux théories des probabilités, il y a un peu moins de 200 ans, il est cocasse de noter qu'aujourd'hui, un petit groupe, non pas d'irréductibles, mais d'actifs chercheurs, travaillent à Gand sous la houlette de Gert de Cooman, à l'expansion et la promotion des théories des probabilités imprécises.

Cette ville semble se prêter aux incertitudes.

Dans cette thèse, nous proposons de remplacer la théorie des probabilités par la théorie des possibilités dans des méthodes de traitement du signal et des statistiques, afin de prendre en compte, de façon plus fiable, les différents défauts sur les données ou sur les méthodes employées.

Nous avons regroupé, dans cette thèse, beaucoup de méthodes de traitement de données (en statistiques ou en traitement du signal) sous un formalisme commun, que nous avons appelé l'**extraction sommative d'informations**. Dans le nom de cette méthode générique, l'expression "extraction d'informations" est explicite : il s'agit de transformer les données en de l'information. Pour ce qui est du terme "sommatif", cela vient du nom de l'outil employé dans les méthodes d'extraction sommative d'informations, que nous nommons le **noyau sommatif**. Le noyau sommatif est une fonction sommant à 1, qui peut être interprétée comme une distribution de probabilité. L'extraction sommative d'informations consiste à regrouper les données dans un voisinage défini par le noyau sommatif, qui joue alors un rôle de fenêtre pondérée, et d'extraire, de ce regroupement, une information par le biais de l'opérateur d'espérance mathématique. C'est une approche intuitive que de regrouper des données pour en extraire de l'information.

Au chapitre I, la définition du formalisme de l'extraction sommative d'informations est accompagnée d'une présentation, dans ce formalisme, de méthodes usuelles de traitement de données. Nous constatons notamment que nous pouvons mettre dans ce cadre générique la modélisation convolutive de l'acquisition d'un signal, la reconstruction d'un signal continu à partir d'un signal échantillonné et l'échantillonnage d'un signal. Ces deux dernières opérations sont à la base de nombreuses applications de traitement numérique du signal et des images telles que les transformations rigides d'une image, le filtrage, la dérivation... En statistiques non paramétriques, beaucoup de méthodes d'estimations fonctionnelles peuvent être incluses dans ce formalisme d'extraction sommative d'informations : les estimateurs de densité de probabilité et de fonction de répartition, que ce soit par histogramme (classique ou flou) ou par la méthode de Parzen Rosenblatt.

Cette unification nous permet de proposer une alternative possibiliste à toutes ces méthodes par le seul remplacement de la méthode d'extraction sommative d'informations. Nous proposons ainsi, au chapitre II, l'utilisation de ce que nous appelons un **noyau maxitif** comme alternative au noyau sommatif. Un noyau maxitif est une fonction dont la borne supérieure est à 1 et qui peut être interprété comme une distribution de possibilité. L'extraction sommative d'informations, qui se faisait par le biais de l'opérateur d'espérance mathématique, est remplacée par l'**extraction maxitive d'informations** qui utilise une généralisation de l'opérateur d'espérance via l'intégrale de Choquet.

Si nous avons choisi la théorie des possibilités, parmi l'ensemble des théories des probabilités imprécises, c'est principalement pour sa simplicité et son applicabilité. Si on travaille sur un domaine à n éléments, la distribution de possibilité nécessite la donnée de n paramètres, alors que les autres modèles de probabilités imprécises, comme la théorie des prévisions cohérentes ou la théorie des fonctions de croyance sont de complexité 2^n. De plus les bornes de l'extraction maxitive d'informations sont obtenues explicitement et directement grâce à l'intégrale de Choquet alors que, dans la plupart des autres théories des probabilité imprécises, des résolutions numériques de programmes linéaires sont nécessaires.

Une différence importante entre une extraction sommative d'informations et une extraction maxitive d'informations consiste en ce que le résultat de cette dernière est un

intervalle, tandis que le résultat de la première est ponctuel. L'intervalle résultant d'une extraction maxitive d'informations reflète l'impact, sur l'information extraite, d'une méconnaissance de l'utilisateur sur le processus modélisé. Dans le chapitre III, nous proposons toute une panoplie de nouvelles méthodes en traitement de signal ou des images et en statistiques. Nous discutons, par exemple, de la modélisation de l'acquisition par extraction maxitive d'informations et de ses liens avec les modèles usuels d'erreur de mesure. Nous proposons une approche maxitive de l'interpolation et de l'échantillonnage d'un signal. Nous proposons aussi une version maxitive de la transformation rigide sur des images. Nous proposons également un filtrage maxitif, qui a pour but de prendre en compte et de quantifier la part de l'arbitraire dans le choix d'un filtre. À partir de cet outil, nous proposons un détecteur de niveau de bruit. Nous avons développé, toujours à partir du filtrage maxitif, un détecteur de contours sur une image, dont la robustesse, vis à vis du bruit, est très nettement améliorée par rapport aux approches usuelles. Nous montrons également que le filtrage maxitif est équivalent à une des extensions floues de la morphologie présentée dans [11]. En statistiques, nous proposons des alternatives robustes au choix du noyau des estimateurs fonctionnels usuels qui ont été présentés au chapitre I.

Chapitre I

Extraction sommative d'informations

Introduction

Dans la plupart des thèses, ce qui suit l'introduction générale, c'est le chapitre étiqueté *état de l'art*. Le but est généralement de placer la problématique et les notions ou outils rencontrés dans la thèse dans l'histoire et la littérature des domaines concernés par nos recherches. Ce chapitre se rapproche quelque peu de cette tradition, mais n'est pas un pur "état de l'art".

Ce que nous proposons dans ce chapitre, c'est de mettre dans un même cadre des méthodes de traitement de données et d'extraction d'informations issues de différents champs applicatifs. Notre but est de formaliser, de poser correctement et rigoureusement la problématique à laquelle nous nous sommes intéressés, afin de rendre nos apports clairs, accessibles et réutilisables.

Nous nous rapprochons ici du principe de l'*état de l'art* au sens où nous allons évoquer des méthodes qui existent déjà dans la littérature. Nous allons notamment présenter les estimateurs à noyau de Parzen Rosenblatt [74, 85] en statistiques ou la modélisation d'un capteur par réponse impulsionnelle en traitement du signal. Nous nous éloignons également du principe d'*état de l'art*, au sens où nous allons, pour toutes ces méthodes, utiliser un même angle d'attaque permettant d'unifier leurs présentations.

Cette présentation unifiée permet de placer toute une batterie de méthodes de traitement de données dans un même cadre : l'**extraction sommative d'informations**. Parmi ces méthodes, on peut citer le filtrage, l'interpolation et l'échantillonnage d'un signal, les estimateurs à noyau de Parzen Rosenblatt de densité de probabilité et de fonction de répartition.

Dans cette présentation unifiée, les données à traiter et les informations à extraire sont prises sous la forme de fonctions synthétisant un ensemble de données ou d'informations ou sous la forme d'un ensemble fini de données ou d'informations discrètes. L'extraction sommative d'informations permet de passer des données aux informations par une transformation s'appuyant sur un outil basique : le **noyau sommatif**.

Le noyau sommatif, sur lequel se base l'extraction sommative d'informations, possède deux interprétations différentes. Soit on est dans une application d'extraction d'informations où la transformation des données en informations est soumise à des incertitudes, alors le noyau sommatif peut être interprété comme une distribution de probabilité. Soit on est dans une application où les données sont pondérées pour être transformées en informations (on a plus d'incertitude dans la transformation et les poids sont les paramètres

de la transformation), alors le noyau sommatif peut être interprété comme une fenêtre pondérée.

Quelle que soit l'interprétation qui en est faite, l'utilisation d'un noyau sommatif est finalement très naturelle pour transformer les données en de l'information. Ce constat explique certainement la grande quantité d'applications que nous avons pu entrer dans le cadre de l'extraction sommative d'informations. La liste d'applications que nous présentons dans ce chapitre n'est certainement pas exhaustive et ne demande qu'à être enrichie.

Nous nous attachons, en section I.1, à présenter, de manière très générale, ce que nous entendons par "extraction d'informations" à partir de données. Nous proposons une présentation rigoureuse des définitions et notations qui seront utilisées dans ce manuscrit. Nous définissons, en section I.2, la méthode d'extraction sommative d'informations et le noyau sommatif ainsi que quelques concepts annexes. Nous discutons également des relations entre l'extraction sommative d'informations et la théorie des distributions de Schwartz [88].

Aux sections I.3 et I.4, nous présentons une boîte à outils de méthodes en traitement du signal et des images puis en statistiques. Toutes ces méthodes sont présentées sous l'angle de l'extraction sommative d'informations. Nous discutons ensuite, en section I.5, d'éventuels indices de comportement des noyaux sommatifs en extraction sommative d'informations. Nous terminons ce chapitre, en section I.6, par une critique de la méthode d'extraction sommative d'informations. Nous proposons enfin une ouverture à l'approche que nous proposerons au chapitre suivant comme alternative à l'extraction sommative d'informations.

I.1 La problématique de l'extraction d'informations

Le cadre théorique dans lequel se place cette thèse, fait de données, d'informations et d'extraction sous forme de fonctions mathématiques, peut paraître a priori peu commun (en particulier les données vues comme des fonctions). Mais nous verrons que beaucoup des méthodes existantes en statistiques et en traitement du signal entrent dans un tel cadre. Ce cadre général est composé de trois entités que sont les données, les informations et l'extraction d'informations à partir de données.

I.1.1 Données

Nous adoptons, tout au long de ce manuscrit, la notation x pour ce qui est des données. x est, d'une manière générale, une fonction :

$$x : \begin{array}{ccc} \Omega & \rightarrow & \mathcal{X}, \\ \omega & \mapsto & x(\omega), \end{array} \tag{I.1}$$

que l'on nomme la **fonction des données**. Ω est l'**ensemble référence des données**. Les données prennent des valeurs dans $\mathcal{X}$, l'**ensemble valeurs des données**, et sont en quelque sorte "indexées" par les éléments de Ω. Une donnée est donc une valeur $x(\omega)$ de $\mathcal{X}$. Un ensemble de données est représenté par l'ensemble des valeurs prises par x sur Ω, c'est-à-dire par l'ensemble $\{x(\omega)|\omega \in \Omega\}$.

On distingue deux cas : les **données continues** et les **données discrètes**. Les données sont dites continues quand Ω est un ensemble infini, et discrètes quand Ω est un ensemble

fini. C'est un abus de langage que nous reprenons au traitement du signal où un signal est dit continu quand son domaine de définition est infini et discret quand son domaine de définition est fini. Il est donc à noter que des données dites continues ne sont pas forcément des fonctions continues au sens classique de l'analyse fonctionnelle.

Dans le cas de données discrètes, l'ensemble référence des données est un ensemble fini $\Omega = \{\omega_1, ..., \omega_n\}$, que l'on peut identifier à une suite d'indices, c'est-à-dire $\Omega := \{1, ..., n\}$. Tout au long de ce manuscrit, l'ensemble de référence Ω fini pour des données discrètes est indifféremment $\{\omega_1, ..., \omega_n\}$ ou $\{1, ..., n\}$. Par cette identification, on peut exprimer la fonction de données x comme une suite de valeurs $(x_i)_{i=1,...,n}$ de $\mathcal{X}$, obtenues pour tout $i \in \{1, ..., n\}$, par :

$$x_i = x(\omega_i).$$

Le cas où les données sont discrètes est le plus fréquent dans les applications courantes de traitement de données. En traitement du signal, il a supplanté le traitement continu lorsque les ordinateurs ont remplacé les processus analogiques. Il est omniprésent dans les applications de statistiques, où l'hypothèse iid est généralement admise pour l'échantillon de données que l'on souhaite traiter :

Définition I.1 (Hypothèse iid) *Soit $(x_i)_{i=1,...,n}$, un ensemble de données. Il est soumis à l'hypothèse iid, si c'est un échantillon de n résultats issus de n expériences régies par n variables aléatoires $X_1, X_2, ..., X_n$ indépendantes (au sens des probabilités) et identiquement distribuées (de même loi de probabilité).*

Actuellement, la plupart des traitements modernes de données sont réalisés de façon discrète. On pourrait légitimement se questionner sur l'intérêt que présente alors le cas continu. Ce cas est particulièrement intéressant dans la pratique, car la grande majorité des problèmes, tant en statistiques qu'en traitement du signal, sont appréhendés dans l'espace continu. En fait le traitement discret des données s'appuie souvent sur un modèle continu sous-jacent et nécessite la modélisation du passage d'un espace discret à un espace continu et vice-versa. Par exemple, dériver un signal échantillonné consiste à estimer la dérivée du signal continu obtenu par interpolation en chaque point d'échantillonnage.

Notons $\mathbb{D}$ l'**ensemble des fonctions de données** associées à un problème. $\mathbb{D}$ peut, par exemple, être $\mathcal{X}^n$, dans le cas où l'on obtient un ensemble de données issues de l'hypothèse iid. $\mathbb{D}$ peut être $C(\Omega, \mathcal{X})$, l'ensemble des fonctions continues de Ω dans $\mathcal{X}$.

Ω peut être fini ou infini, et on a des données discrètes ou continues. L'ensemble $\mathcal{X}$, des valeurs que peuvent prendre les données, peut lui aussi être fini ou infini. Dans le cas où $\mathcal{X}$ est fini, on parle de **données quantifiées** (pour $\mathcal{X}$ infini, on ne dit rien d'autre que des données non quantifiées). On peut, par exemple, avoir des données quantifiées issues d'un lancé de dés pouvant prendre des valeurs dans $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ ou des données non quantifiées prise par un thermomètre dont l'éventail des valeurs est $\mathcal{X} = [-50, 100]$.

En général, on parle de **données échantillonnées** pour des données discrètes. Pour des données à la fois échantillonnées et quantifiées (i.e. Ω et $\mathcal{X}$ finis), on parle de **données numériques** ou de **données numérisées**. C'est le type de données traitées par ordinateur. Dans ce manuscrit, ces termes seront employés sans distinction car la quantification des données n'intervient pas dans les résultats que nous présentons.

Voici un petit tableau récapitulatif des notations que nous utilisons :

Notation	Définition
Ω	ensemble référence des données
$\mathcal{X}$	ensemble valeurs des données
$x : \Omega \to \mathcal{X}$	fonction de données
$\mathbb{D}$	ensemble des fonctions de données
données continues :	Ω infini
$x(\omega)$	une donnée
$\{x(\omega)\|\omega \in \Omega\}$	ensemble des données
données discrètes :	$\Omega = \{1, ..., n\}$
x_i	une donnée
$(x_i)_{i=1,...,n}$	ensemble des données

Tab. I.1 : Notations des donneées

I.1.2 Informations

Le modèle présenté ici pour les informations est analogue au modèle des données de la section précédente. Seules les notations et les dénominations sont différentes. Les informations sont également modélisées par une fonction. Nous adoptons la notation y pour la **fonction d'informations** :

$$y : \begin{array}{ccc} \Theta & \to & \mathcal{Y}, \\ \theta & \mapsto & y(\theta). \end{array} \tag{I.2}$$

L'ensemble Θ est l'**ensemble référence des informations**. $\mathcal{Y}$, l'**ensemble valeurs des informations**, est l'ensemble des valeurs que les informations peuvent prendre. Une information est donc une valeur $y(\theta)$ de $\mathcal{Y}$. Un ensemble d'informations est un ensemble $\{y(\theta)|\theta \in \Theta\}$.

On distingue également deux cas : les **informations continues** et les **informations discrètes**. Les informations sont dites continues quand Θ est un ensemble infini et discrètes quand Θ est un ensemble fini. Dans le cas d'informations discrètes, l'ensemble référence des informations est un ensemble fini $\Theta = \{\theta_1, ..., \theta_p\}$, que l'on peut identifier à une suite d'indices. Ainsi dans le cas où les informations sont discrètes, $\Theta := \{1, ..., p\}$. Tout au long de ce manuscrit, l'ensemble de référence Θ fini pour des informations discrètes est indifféremment $\{\theta_1, ..., \theta_p\}$ ou $\{1, ..., p\}$. Par cette identification, on peut exprimer la fonction d'informations y comme une suite de valeurs $(y_k)_{k=1,...,p}$ de $\mathcal{Y}$, obtenues pour tout $k \in \{1, ..., p\}$, par :

$$y_k = y(\theta_k). \tag{I.3}$$

On note $\mathbb{I}$ l'**ensemble des fonctions d'informations**. $\mathbb{I}$ peut, par exemple, être $\mathcal{Y}^p$, pour des cas d'estimations de p paramètres d'une loi à partir d'un échantillon statistique iid. On peut avoir $\mathbb{I} = \{y \in L_1(\Theta)| \int_\Theta y(\theta)d\theta = 1\}$. C'est le cas de l'estimation de densité de probabilité.

L'ensemble $\mathcal{Y}$, des valeurs que peuvent prendre les informations, peut être également discret. On parlera alors d'**informations quantifiées**.

En général, on parle d'**informations échantillonnées** pour des informations discrètes. Pour des informations à la fois discrètes et quantifiées (i.e. Θ et $\mathcal{Y}$ finis), on parle d'**informations numériques** ou d'**informations numérisées**. Dans ce manuscrit, ces

termes seront employés sans distinction car la quantification des informations n'intervient pas dans les résultats que nous présentons.

Voici un petit tableau récapitulatif des notations que nous utilisons :

Notation	Définition
Θ	ensemble référence des informations
$\mathcal{Y}$	ensemble valeurs des informations
$y : \Theta \to \mathcal{Y}$	fonction de informations
$\mathbb{I}$	ensemble des fonctions de informations
informations continues :	Θ infini
$y(\theta)$	une information
$\{y(\theta)\|\theta \in \Theta\}$	ensemble des informations
informations discrètes :	$\Theta = \{1, ..., p\}$
y_k	une information
$(y_k)_{k=1,...,n}$	ensemble des informations

Tab. I.2 : Notations des informations

I.1.3 Extraction précise d'informations

De manière intuitive, extraire des informations à partir d'une fonction de données x de $\mathbb{D}$, c'est transformer cette fonction de données en une fonction d'informations, c'est-à-dire en une fonction y de $\mathbb{I}$. On parle d'extraction précise d'informations, car on extrait une unique fonction d'informations. On peut formaliser mathématiquement l'extraction précise d'informations par la donnée d'une fonction de transformation h :

$$h : \begin{array}{ccc} \mathbb{D} & \to & \mathbb{I}, \\ x & \mapsto & y = h(x), \end{array} \tag{I.4}$$

Cette extraction est réalisée de manière globale. Le résultat de l'extraction $h(x)$ est une fonction d'informations définie pour tout θ de Θ. La figure I.1 présente un exemple qualitatif et arbitraire d'extraction précise globale d'informations.

L'extraction locale d'information revient à transformer une fonction de données x en une valeur de l'information $y(\theta)$ en θ. L'extraction locale d'information en θ peut donc être résumée par la donnée d'une fonction h^θ qui transforme un ensemble de données $\{x(\omega)|\omega \in \Omega\}$ en une information $y(\theta)$ en θ :

$$h^\theta : \begin{array}{ccc} \mathbb{D} & \to & \mathcal{Y}, \\ x & \mapsto & y(\theta) = h^\theta(x), \end{array} \tag{I.5}$$

La donnée de la fonction h^θ, pour obtenir un élément d'information $y(\theta)$, renseigne sur la manière dont sont transformées les données $\{x(\omega)|\omega \in \Omega\}$ en $y(\theta)$. La figure I.1 présente un exemple qualitatif et arbitraire d'extraction précise locale d'information en θ.

Extraire globalement de l'information revient à définir, pour tout θ de Θ, sa fonction d'extraction locale h^θ. Pour une fonction de données quelconque $x \in \mathbb{D}$, $h(x)(\theta) = h^\theta(x)$, $\forall \theta \in \Theta$.

Définir les transformations $(h^\theta)_{\theta\in\Theta}$, c'est-à-dire définir la transformation globale h, c'est définir la nature des informations que l'on souhaite extraire. Par exemple, transformer

les données en leur moyenne revient à extraire l'information de moyenne des données. Dans cet exemple on a une seule information à extraire : la moyenne. Donc Θ est un ensemble qui ne contient qu'un élément. Prenons maintenant l'exemple du filtrage par filtre passe bas d'un signal : transformer un signal en un autre signal ne contenant que les parties de basses fréquences d'un signal (filtre passe bas), revient à extraire comme informations les parties de basses fréquences du signal, ce qui revient souvent à réaliser des moyennes locales.

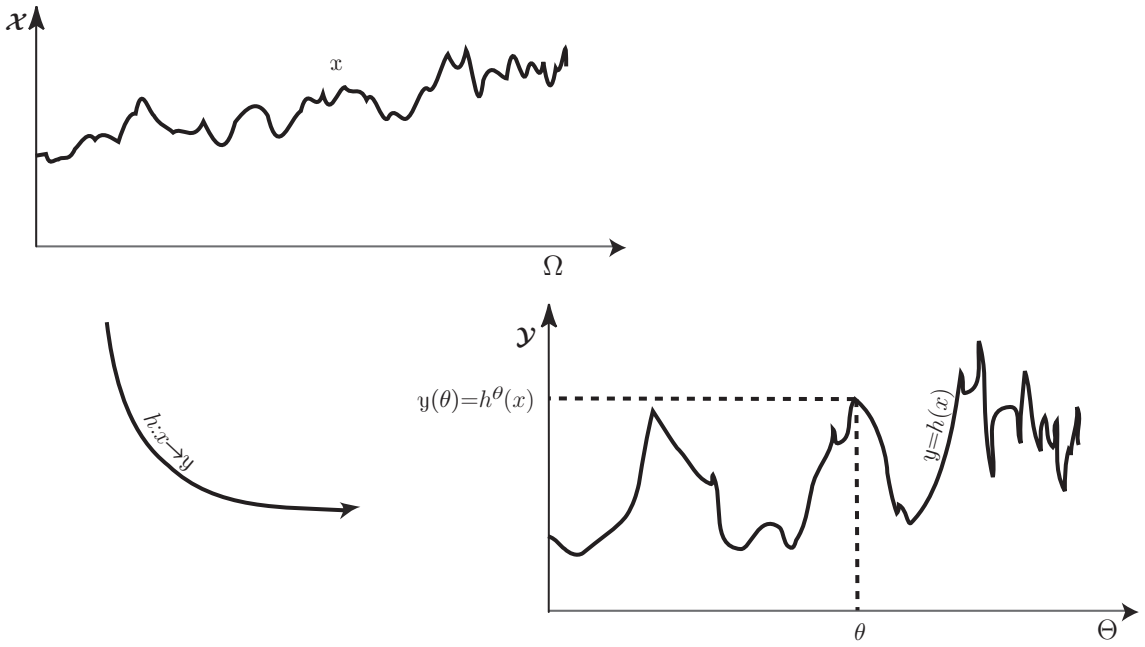


Fig. I.1 : Schéma général de l'extraction d'information

La figure I.1 se limite au cas où les données et les informations à extraire sont continues. La nature des données à traiter et de l'information à extraire ne se bornent pas à cette représentation continue. On différencie quatre cas qui correspondent à des combinaisons de données discrètes ou continues et des informations discrètes ou continues. Il faut pour chacun de ces quatre cas, adapter la nature de la fonction d'extraction locale h^θ aux types des données et des informations.

Dans beaucoup d'applications, la fonction de données et la fonction d'informations ont le même domaine de référence. On retrouve cela quand on fait, par exemple, du filtrage de signal. Dans ce cas, $\Omega = \Theta$, et on a $x : \Omega \to X$ et $y : \Omega \to \mathcal{Y}$.

I.2 Extraction sommative d'informations

Nous avons proposé un problème très général. Le but de cette section est de le particulariser afin d'en ressortir un modèle directement manipulable et utilisable. Nous nous restreignons à des données et des informations à valeurs réelles, i.e. $\mathcal{X} = \mathcal{Y} = \mathbb{R}$.

La clef de ce nouveau modèle ne tient pas dans cette restriction aux réels, mais sur l'ajout d'une hypothèse sur la transformation des données en une information, i.e. sur l'extraction locale d'informations. Ce nouveau modèle est nommé *l'extraction sommative d'informations* car il s'appuie sur l'hypothèse que les données sont réparties suivant une

distribution connue en tout point de la fonction d'informations, et représentable par un noyau sommatif.

I.2.1 La clef : le noyau sommatif

Définition I.2. *Un noyau sommatif κ est une fonction définie sur Ω à valeur dans $\mathbb{R}^+$, vérifiant la propriété de sommativité :*

$$\int_\Omega \kappa(\omega)d\omega = 1, \text{ dans le cas continu,} \tag{I.6}$$

$$\sum_{i\in\Omega} \kappa_i = 1, \text{ dans le cas discret.} \tag{I.7}$$

Dans le cas discret, un noyau sommatif peut être considéré comme une suite $(\kappa_i)_{i=1,...,n}$ d'éléments de $\mathbb{R}^+$ ou un vecteur de $\mathbb{R}^{n+}$.

Un noyau sommatif κ peut être vu comme un voisinage pondéré (ou fenêtre pondérée) que l'on place sur le domaine Ω de la fonction de données, et dont la magnitude $\kappa(\omega)$ correspond au poids que l'on attribue à l'élément ω de Ω.

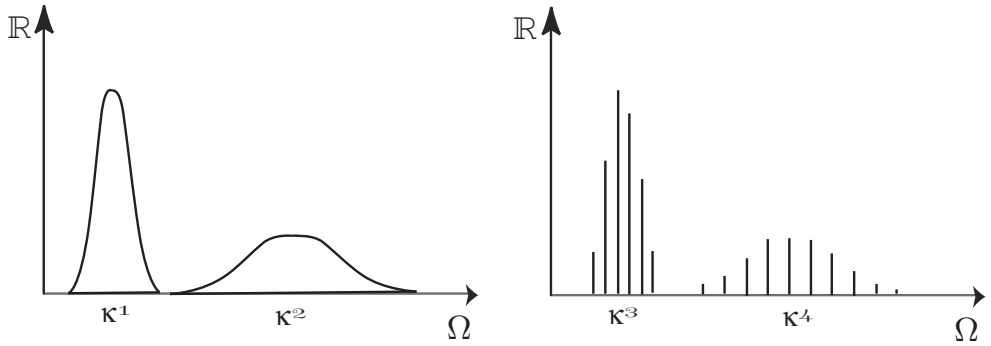


Fig. I.2 : Représentations qualitatives de noyaux sommatifs

Un noyau sommatif peut également être interprété comme une distribution de probabilité. La définition d'un noyau sommatif coïncide parfaitement avec la définition d'une distribution de probabilité. Vu sous cet angle, un noyau sommatif κ définit une mesure de probabilité, notée P_κ, obtenue par :

$$\forall A \subseteq \Omega,\ P_\kappa(A) = \int_A \kappa(\omega)d\omega, \text{ dans le cas continu,} \tag{I.8}$$

$$\forall A \subseteq \Omega = \{1, ..., n\},\ P_\kappa(A) = \sum_{i\in A} \kappa_i, \text{ dans le cas discret,} \tag{I.9}$$

vérifiant l'axiome d'additivité de Kolmogorov, disant que, pour toute suite dénombrable d'évènements deux à deux disjoints $(A_m)_{m>0} \subseteq \Omega$, on a :

$$P_\kappa(\bigcup_{m>0} A_m) = \sum_{m>0} P_\kappa(A_m). \tag{I.10}$$

Dans le cas discret, l'axiome suivant est suffisant :

$$\forall A, B \subseteq \Omega, \text{ disjoints, } P_\kappa(A \cup B) = P_\kappa(A) + P_\kappa(B). \tag{I.11}$$

Notons $\mathbb{S}_c(\Omega)$, l'ensemble des noyaux sommatifs continus sur Ω.

Notons $\mathbb{S}_d(\Omega)$, l'ensemble des noyaux sommatifs discrets sur Ω.

Notons $\mathbb{S}(\Omega)$, la réunion de ces ensembles, c'est-à-dire l'ensemble des noyaux sommatifs sur Ω.

I.2.2 L'extraction sommative d'informations

Dans la partie I.1.3, nous avons introduit la notion d'extraction locale précise d'information sous la forme d'une fonction h^θ transformant la fonction de données x en une information $y(\theta)$. Cette fonction h^θ est le cœur de l'extraction précise d'informations. Dans la méthode d'extraction sommative d'information, c'est un noyau sommatif κ^θ qui caractérise la façon dont sont agrégées les données $\{x(\omega)|\omega \in \Omega\}$ pour obtenir la valeur de l'information $y(\theta)$. Du choix du noyau, dépend l'information que l'on souhaite extraire.

Le noyau sommatif κ^θ est défini sur Ω, l'ensemble référence des données. À chaque élément ω de Ω, correspondent une donnée $x(\omega)$ et un poids $\kappa^\theta(\omega)$ attribué à $x(\omega)$ pour l'extraction d'information locale $y(\theta)$. Dans ce cadre restreint de l'extraction sommative d'informations, l'extraction proprement dite est réalisée par un opérateur d'agrégation [65]. Encore une fois, il est important de distinguer le cas où les données sont continues du cas où les données sont discrètes.

Dans le cas où les données sont continues, c'est le produit scalaire dans $L_1(\Omega) = \{f : \Omega \to \mathbb{R} \mid \int_\Omega f(\omega)d\omega < +\infty\}$ entre la fonction de données continue x et le noyau sommatif continu κ^θ qui est utilisé comme opérateur d'agrégation. Pour ce qui est du noyau sommatif κ^θ, il appartient bien à l'ensemble des fonctions intégrables $L_1(\Omega)$, car un noyau sommatif intègre à 1 d'après la propriété de sommativité (I.6). Mais cette opérateur d'agrégation impose à la fonction de données d'être également une fonction de $L_1(\Omega)$.

L'extraction sommative d'information en θ à partir de données continues est obtenue par :

$$y(\theta) = \left\langle x, \kappa^\theta \right\rangle_{L_1(\Omega)} = \int_\Omega x(\omega)\kappa^\theta(\omega)d\omega, \tag{I.12}$$

pour $x \in L_1(\Omega)$ et $\kappa^\theta \in \mathbb{S}_c(\Omega)$.

Dans le cas discret, les données et x et le noyau sommatif sont des vecteurs de $\mathbb{R}^n$. L'opérateur d'agrégation utilisé est le produit de scalaire dans $\mathbb{R}^n$ ou dans $l_1(\Omega) = l_1(\{1, ..., n\} = \{u = (u_i)_{i=1,...,n} \subseteq \mathbb{R} | \ \sum_{i=1}^n u_i < +\infty\}$ entre le vecteur de données x et le vecteur du noyau sommatif κ^θ. Le produit scalaire dans $\mathbb{R}^n$ est défini pour $u, v \in l_1(\Omega)$, $\langle u, v\rangle_{l_1(\Omega)} = \sum_{i=1}^n u_i v_i$. Cette opérateur d'agrégation impose à la fonction de données x d'être également une fonction de $l_1(\Omega)$.

L'extraction sommative d'information en θ à partir de données discrètes est obtenue par :

$$y(\theta) = \left\langle x, \kappa^\theta \right\rangle_{l_1(\Omega)} = \sum_{i=1}^n x_i \kappa_i^\theta, \tag{I.13}$$

pour $x \in l_1(\Omega)$ et $\kappa^\theta \in \mathbb{S}_d(\Omega)$.

Les expressions (I.12) et (I.13) étant très similaires, nous proposons de regrouper les opérateurs d'agrégation que sont les produits scalaires dans $l_1(\Omega)$ et dans $L_1(\Omega)$ et ainsi définir un seul et même opérateur d'agrégation, l'espérance mathématique pour le noyau sommatif κ^θ.

Définition I.3 (Extraction sommative d'information)
Sous les hypothèses :

1. $x \in l_1(\Omega)$ *(resp.* $L_1(\Omega)$*),*
2. $\kappa^\theta \in \mathbb{S}_c(\Omega)$ *(resp.* $\mathbb{S}_d(\Omega)$*),*

l'extraction sommative d'information en θ *est obtenue par*

$$y(\theta) = \mathbb{E}_{\kappa^\theta}(x). \tag{I.14}$$

Autrement dit, la fonction d'extraction h^θ en θ est l'espérance mathématique de la fonction de données x suivant la loi probabiliste associée à κ^θ, i.e. $h^\theta = \mathbb{E}_{\kappa^\theta}$.

Dans le cas où les données sont continues, les noyaux sommatifs utilisés pour obtenir de l'information sur Θ sont continus. Dans le cas où les données sont discrètes, les noyaux sommatifs utilisés pour obtenir de l'information sur Θ sont discrets. La nature de l'information importe peu.

Si l'information à extraire est continue, l'obtention de $y(\theta)$, pour tout θ de Θ, nécessite une infinité de noyaux sommatifs (discrets ou continus). Si la fonction d'informations est discrète, seul un nombre fini de noyaux sommatifs (discrets ou continus) est nécessaire pour extraire de façon globale l'information.

Dans le cas où la fonction de données x et celle d'informations y ont le même ensemble référence Ω continu, on utilise pour chaque ω de Ω, un noyau sommatif $\kappa^\omega : \Omega \to \mathbb{R}$. L'extraction sommative d'informations continue sur les mêmes ensembles référence devient alors :

$$y(\omega) = \int_\Omega x(u)\kappa^\omega(u)du. \tag{I.15}$$

Dans le cas où la fonction de données x et celle d'informations y ont le même ensemble référence Ω discret, on utilise alors pour chaque k de Ω, un noyau sommatif $\kappa^k : \Omega \to \mathbb{R}$. L'extraction sommative d'informations discrète sur les mêmes ensembles référence devient alors :

$$y_k = \sum_{i=1}^{n} x_i \kappa_i^k. \tag{I.16}$$

Dans cette approche où $\Theta = \Omega$, à chaque valeur ω de Ω est associé un noyau sommatif κ^ω. Dans la plupart des applications utilisant l'extraction sommative d'informations, ce noyau sommatif κ^ω est supposé invariant par translation, i.e. il existe un noyau sommatif de référence κ, tel que :

$$\forall u \in \Omega,\ \kappa^\omega(u) = \kappa(\omega - u). \tag{I.17}$$

Définir un noyau sommatif de référence (I.17) invariant par translation en tout point de Ω, nécessite que l'espace Ω soit stable par soustraction. Si ce n'est pas le cas, on peut étendre Ω en un autre espace, $\overline{\Omega}$, reconstruit de telle sorte qu'il soit stable par soustraction.

Par exemple, dans le cas où Ω est un sous - ensemble des réels, on peut étendre Ω à $\overline{\Omega} = \mathbb{R}$, qui lui, est stable par soustraction. Par convention, les valeurs des poids $\kappa(v)$ et des données $x(v)$, pour tout v dans $\mathbb{R}\backslash\Omega$, sont nulles.

Dans le cas discret, Ω est toujours identifié à l'ensemble des indices $\{1, ..., n\}$. On peut donc toujours l'étendre à l'ensemble des entiers relatifs $\mathbb{Z}$ qui est stable par soustraction. Les valeurs des poids κ_m et des données x_m, pour tout m dans $\mathbb{Z}\backslash\Omega$, sont, par convention, nulles.

Dans le cas discret, sous l'hypothèse que l'on a un noyau sommatif κ de référence invariant par translation, κ^k, le translaté de κ en $k \in \Omega$ est obtenu par :

$$\forall i \in \mathbb{Z},\ \kappa_i^k = \kappa_{k-i}. \tag{I.18}$$

L'extraction sommative d'information devient alors :

$$y(\omega) = \int_{\overline{\Omega}} x(u)\kappa(\omega - u)du,\ \text{dans le cas continu,} \tag{I.19}$$

$$y_k = \sum_{i\in\mathbb{Z}} x_i\kappa_{k-i},\ \text{dans le cas discret.} \tag{I.20}$$

Dans ce contexte particulier où l'agrégation en tout point de l'ensemble référence d'informations est régie par un seul noyau sommatif, l'information à extraire est le produit de convolution de la fonction de données avec ce noyau sommatif. On appellera alors κ le noyau sommatif de convolution.

L'extraction sommative d'informations peut donc s'écrire, respectivement dans le cas continu et discret, comme :

$$y(\omega) = (x * \kappa)(\omega) \text{ et } y_k = (x * \kappa)_k. \tag{I.21}$$

Cette approche convolutive est un cas particulier de l'extraction sommative d'informations sur deux mêmes ensembles de référence. Il nous semble important de bien insister sur ce fait : l'extraction sommative d'informations à partir de données ayant les mêmes ensembles référence Ω, n'implique pas l'utilisation de la convolution. L'utilisation d'un seul et même noyau de référence pour régir l'agrégation de l'information en tout point de Ω nécessite souvent l'utilisation d'hypothèses fortes justifiant l'invariance par translation lors du traitement.

On pourra également remarquer que chaque noyau sommatif continu, défini sur $\mathbb{R}$ peut être à l'origine d'une famille de noyaux sommatifs paramétrée par la largeur de bande $\Delta > 0$. Chaque élément de cette famille peut être obtenu par :

$$\kappa_\Delta(\omega) = \frac{1}{\Delta}\kappa(\frac{\omega}{\Delta}),\ \forall\omega \in \Omega. \tag{I.22}$$

Voici par exemple cinq éléments de la famille de noyaux d'Epanechnikov, obtenus par $\forall\omega \in \Omega,\ \kappa_\Delta(\omega) = \frac{3}{4\Delta}(1 - \frac{\omega^2}{\Delta^2})\mathbb{1}_{|\omega|\leq\Delta}$.

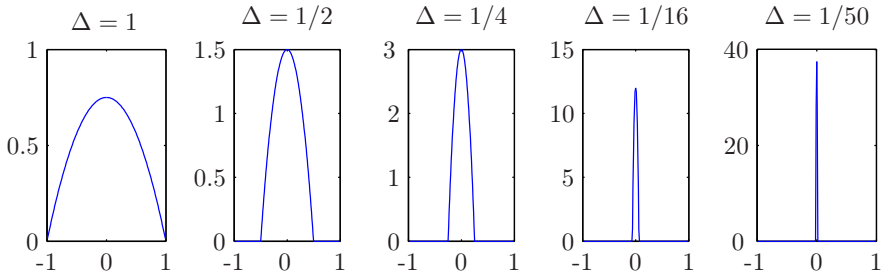


Fig. I.3 : Famille de noyaux sommatifs d'Epanechnikov

La largeur de bande est un facteur de dilatation du noyau κ_Δ.

I.2.3 Extraction sommative séquentielle d'informations

L'extraction sommative d'informations, définie à la section précédente, s'avère utile lorsque l'on est capable de choisir un noyau sommatif qui régit la répartition des données sur la fonction d'informations. Il est parfois plus simple de procéder par étapes pour extraire des informations. Effectuer une suite d'extractions sommatives d'informations pour mener à de l'information finale peut s'avérer plus réalisable que de transformer directement les données de départ en cette information finale. Une propriété importante de l'extraction sommative d'informations est résumée par le théorème suivant :

Théorème I.4 (Extraction sommative séquentielle d'informations)
Une séquence d'extraction sommatives d'informations est équivalente à une extraction sommative d'informations.

Preuve : Voyons cela pour une séquence de deux extractions sommatives d'informations consécutives. Dans un premier temps, on extrait de l'information y à partir des données de départ x avec les noyaux sommatifs κ^θ, définis pour tous θ de Θ. On a donc, d'après l'expression (I.14) que $\forall\theta \in \Theta$, $y(\theta) = \mathbb{E}_{\kappa^\theta}(x)$. Si l'on extrait maintenant l'information z, que l'on formalise par une fonction

$$z : \begin{array}{ccc} \Xi & \rightarrow & \mathcal{Z}, \\ \xi & \mapsto & z(\xi), \end{array} \tag{I.23}$$

à partir des données intermédiaires y, on a donc, d'après l'expression (I.14) que $\forall\xi \in \Xi$, $z(\xi) = \mathbb{E}_{\kappa^\xi}(y)$.

Soit η^ξ, un noyau défini par :

$$\eta^\xi : \begin{array}{ccc} \Omega & \rightarrow & \mathbb{R}, \\ \omega & \mapsto & \eta^\xi(\omega) = \int_\Theta \kappa^\theta(\omega)\kappa^\xi(\theta)d\theta. \end{array} \tag{I.24}$$

η^ξ est un noyau sommatif. En effet,

$$\begin{aligned} \int_\Omega \eta^\xi(\omega)d\omega &= \int_\Omega \int_\Theta \kappa^\theta(\omega)\kappa^\xi(\theta)d\theta d\omega, \\ &= \int_\Theta \kappa^\xi(\theta) \int_\Omega \kappa^\theta(\omega)d\omega d\theta, \\ &= \int_\Theta \kappa^\xi(\theta)d\theta = 1. \end{aligned}$$

Montrons maintenant que $z(\xi)$, l'information finale en ξ, est obtenu par extraction sommative d'information, avec le noyau sommatif η^ξ, à partir des données de départ x. $\forall \xi \in \Xi$,

$$\begin{aligned}
z(\xi) &= \mathbb{E}_{\kappa^\xi}(y), \\
&= \mathbb{E}_{\kappa^\xi}(\mathbb{E}_{\kappa^\theta}(x)), \\
&= \int_\Theta \mathbb{E}_{\kappa^\theta}(x)\kappa^\xi(\theta)d\theta, \\
&= \int_\Theta \big(\int_\Omega x(\omega)\kappa^\theta(\omega)d\omega\big)\kappa^\xi(\theta)d\theta, \\
&= \int_\Theta \int_\Omega x(\omega)\kappa^\theta(\omega)\kappa^\xi(\theta)d\omega d\theta, \text{ par Fubini} \\
&= \int_\Omega \int_\Theta x(\omega)\kappa^\theta(\omega)\kappa^\xi(\theta)d\theta d\omega, \\
&= \int_\Omega x(\omega)\big(\int_\Theta \kappa^\theta(\omega)\kappa^\xi(\theta)d\theta\big)d\omega, \\
&= \int_\Omega x(\omega)\eta^\xi(\omega)d\omega, \\
&= \mathbb{E}_{\eta^\xi}(x).
\end{aligned}$$

L'extension de la preuve à une suite d'extractions sommatives d'informations avec plus de deux extractions est immédiate. De même, la preuve en discret est triviale.

□

I.2.4 Extraction sommative séparable d'informations

Dans beaucoup d'applications, les données à traiter sont définies sur des ensembles référence Ω qui sont des produits cartésiens d'ensembles. C'est le cas des données multidimensionnelles. Par exemple, en traitement analogique (continu) d'images, Ω est un sous-ensemble compact de $\mathbb{R}^2$.

Dans la plupart des applications basées sur des données multidimensionnelles, on utilise, pour simplifier les calculs, des noyaux sommatifs dits séparables.

Définition I.5. *Le noyau sommatif κ_{12}, défini sur $\Omega = \Omega_1 \times \Omega_2$, est dit séparable, quand il peut s'exprimer comme le produit des noyaux sommatifs κ_1 sur Ω_1 et κ_2 sur Ω_2 :*

$$\forall(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2,\ \kappa_{12}(\omega_1, \omega_2) = \kappa_1(\omega_1)\kappa_2(\omega_2). \tag{I.25}$$

Théorème I.6 (Extraction sommative d'informations séparable)
Soit κ_{12}^θ un noyau sommatif séparable défini sur $\Omega = \Omega_1 \times \Omega_2$. L'extraction sommative d'informations à partir d'une fonction de données x définie sur $\Omega = \Omega_1 \times \Omega_2$ est équivalente à une extraction sommative séquentielle d'informations faisant appel (dans n'importe quel ordre) aux noyaux sommatifs κ_1^θ et κ_2^θ.

Preuve : Posons $x(\bullet, \omega_2)$, la fonction mono dimensionnelle, à ω_2 fixée, définie pour tout $\omega_1 \in \Omega_1$, par $x(\bullet, \omega_2)(\omega_1) = x(\omega_1, \omega_2)$. On note $y_1^\theta(\omega_2)$, l'information extraite en $\theta \in \Theta$, à partir de $x(\bullet, \omega_2)$ avec le noyau sommatif κ_1^θ. Autrement dit, $y_1^\theta(\omega_2) = \mathbb{E}_{\kappa_1^\theta}(x(\bullet, \omega_2))$, pour tout $\omega_2 \in \Omega_2$.

Soit $y(\theta)$ l'information extraite avec le noyau sommatif κ_{12}^{θ} en $\theta \in \Theta$. on a :

$$\begin{aligned} y(\theta) &= \mathbb{E}_{\kappa_{12}^{\theta}}(x), \\ &= \int_{\Omega_2}\int_{\Omega_1} x(\omega_1,\omega_2)\kappa_1^{\theta}(\omega_1)\kappa_2^{\theta}(\omega_2)d\omega_1 d\omega_2, \\ &= \int_{\Omega_2} \mathbb{E}_{\kappa_1^{\theta}}(x(\bullet,\omega_2))\kappa_2^{\theta}(\omega_2)d\omega_2, \\ &= \int_{\Omega_2} y_1^{\theta}(\omega_2)\kappa_2^{\theta}(\omega_2)d\omega_2, \\ &= \mathbb{E}_{\kappa_2^{\theta}}(y_1^{\theta}). \end{aligned}$$

Posons $x(\omega_1,\bullet)$, la fonction mono dimensionnelle, à ω_1 fixée, définie pour tout $\omega_2 \in \Omega_2$, par $x(\omega_1,\bullet)(\omega_2) = x(\omega_1,\omega_2)$. On note $y_2^{\theta}(\omega_1)$, l'information extraite en $\theta \in \Theta$, à partir de $x(\omega_1,\bullet)$ avec le noyau sommatif κ_2^{θ}. Autrement dit, $y_2^{\theta}(\omega_1) = \mathbb{E}_{\kappa_2^{\theta}}(x(\omega_1,\bullet))$, pour tout $\omega_2 \in \Omega_2$.

$$\begin{aligned} y(\theta) &= \mathbb{E}_{\kappa_{12}^{\theta}}(x), \\ &= \int_{\Omega_1}\int_{\Omega_2} x(\omega_1,\omega_2)\kappa_1^{\theta}(\omega_1)\kappa_2^{\theta}(\omega_2)d\omega_2 d\omega_1, \\ &= \int_{\Omega_1} \mathbb{E}_{\kappa_2^{\theta}}(x(\omega_1,\bullet))\kappa_1^{\theta}(\omega_1)d\omega_1, \\ &= \int_{\Omega_1} y_2^{\theta}(\omega_1)\kappa_1^{\theta}(\omega_1)d\omega_1, \\ &= \mathbb{E}_{\kappa_1^{\theta}}(y_2^{\theta}). \end{aligned}$$

□

I.2.5 Extraction sommative et distribution de Schwartz

Le but de cette partie est de discuter du lien entre l'extraction sommative continue d'informations et la théorie des distributions de Schwartz [88]. Le précédent détour par la convolution a déjà pu te mettre, Lecteur, sur la voie des distributions. Mais c'est surtout la formulation sous forme de produit scalaire de $L_1(\Omega)$ qui lie cette méthodologie sommative à la théorie des distributions. La théorie des opérateurs (ou transformations) intégrales [79] est également évoquée ici.

L'idée de base de la théorie des distributions peut être énoncée comme suit[1] :

> On évalue habituellement une fonction en calculant sa valeur en un point. Toutefois cette méthode fait jouer un rôle considérable aux irrégularités (discontinuités par exemple) de la fonction. L'idée sous-jacente à la théorie des distributions est qu'il existe un meilleur procédé d'évaluation : calculer une moyenne des valeurs de la fonction dans un domaine de plus en plus resserré

[1] source Wikipedia [112]

autour du point d'étude. En envisageant des moyennes pondérées, on est donc conduit à examiner des expressions de la forme

$$T_f(\varphi) = \int_\Omega f(\omega)\varphi(\omega)d\omega, \tag{I.26}$$

dans laquelle la fonction à évaluer $f : \Omega \to \mathbb{R}$ est une fonction localement intégrable et $\varphi : \Omega \to \mathbb{R}$ est une fonction appelée "fonction test".

D'une manière plus générale, une distribution est un objet qui généralise la notion de fonction. Les distributions sont des fonctionnelles à valeurs réelles, c'est-à-dire des fonctions de fonctions. Dans la théorie de Schwartz [88], les distributions ont pour arguments les "fonctions test" qui sont des fonctions à valeurs réelles infiniment dérivables et à support compact. L'ensemble des fonctions test sur Ω est noté $\mathcal{D}(\Omega)$. L'espace des distributions est donc le dual topologique de $\mathcal{D}(\Omega)$, c'est-à-dire l'ensemble des formes linéaires sur $\mathcal{D}(\Omega)$. L'espace des distributions est alors $\mathcal{D}'(\Omega) = \{T : \mathcal{D}(\Omega) \to \mathbb{R}\}$. L'introduction aux distributions de Schwartz, présentée au-dessus, ne parle que d'un cas particulier de distributions : les distributions T_f que l'on peut identifier avec des fonctions f localement intégrables. Cette introduction a le mérite de mettre en évidence l'analogie qui existe entre la théorie des distributions et l'extraction sommative d'information.

Dans le cas de la théorie des distributions, le calcul de la distribution s'appuie sur le produit scalaire, dans $L_1(\Omega)$, de la fonction f avec le voisinage qu'est la fonction test φ. D'après l'expression (I.26), $T_f(\varphi) = \langle f, \varphi \rangle_{L_1(\Omega)}$. Par abus de langage, on identifie souvent la distribution I_f à la fonction f.

Dans le cas de l'extraction sommative d'information, le calcul de l'information $y(\theta)$ en θ, consiste à calculer le produit scalaire de la fonction de données x (dont on cherche à extraire de l'information) avec le voisinage qu'est le noyau sommatif κ^θ en θ. En effet, d'après l'expression (I.12), $y(\theta) = \left\langle x, \kappa^\theta \right\rangle_{L_1(\Omega)}$.

Dans les deux cas, distribution ou extraction sommative d'informations, une hypothèse d'intégrabilité est imposée aux fonctions f ou x à traiter. Pour ce qui est des fonctions test et des noyaux sommatifs, on n'a pas l'inclusion de l'ensemble des fonctions test dans l'ensemble des noyaux sommatifs ni inversement. En effet, une fonction test $\varphi \in \mathcal{D}(\Omega)$ n'est pas forcément sommative, i.e. $\int_\Omega \varphi(\omega)d\omega \neq 1$. Donc $\mathcal{D}(\Omega) \not\subseteq \mathbb{S}_c(\Omega)$. De même un noyau sommatif continu $\kappa \in \mathbb{S}_c(\Omega)$ n'est pas forcément infiniment différentiable, donc $\mathbb{S}_c(\Omega) \not\subseteq \mathcal{D}(\Omega)$.

Il existe cependant une intersection entre ces deux ensembles de fonctions (noté $\mathbb{S}_c(\Omega) \cap \mathcal{D}(\Omega)$) qui sont les fonctions test vérifiant la propriété de sommativité ou les noyaux sommatifs infiniment différentiables (donc continus) à support borné.

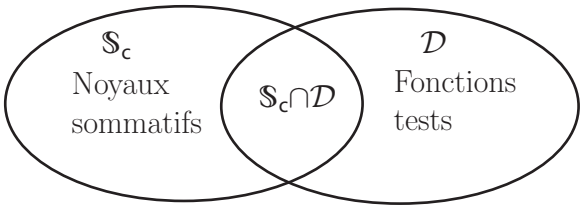


Fig. I.4 : Noyaux sommatifs et fonctions test

Il en résulte une forte analogie entre le calcul de distribution et l'extraction sommative d'informations. Cependant, aucun des deux n'est un cas particulier de l'autre.

L'intérêt de la condition de différentiabilité infinie des fonctions test tient au fait que les dérivées (jusqu'à l'infini) d'une distribution sont obtenues grâce à la propriété suivante :

$$\forall k \in \mathbb{N},\ d^{(k)}T_f(\varphi) = \left\langle d^{(k)}f, \varphi \right\rangle_{L_1(\Omega)} = (-1)^k \left\langle f, d^{(k)}\varphi \right\rangle_{L_1(\Omega)}. \tag{I.27}$$

La condition de support borné intervient dans la démonstration de la seconde égalité de l'expression (I.27).

D'un point de vue théorique, il est nécessaire de définir la dérivée d'une distribution pour une infinité de rangs. Mais dans notre cadre applicatif, qu'est l'extraction sommative d'informations, définir cette dérivée pour une infinité de rangs ne paraît pas toujours utile. En effet, si l'on définit la dérivation de l'extraction sommative d'informations de la même manière, i.e.

$$\textit{pour } k \in \mathbb{N},\ d^{(k)}y(\theta) = \left\langle d^{(k)}x, \kappa^\theta \right\rangle_{L_1(\Omega)} = (-1)^k \left\langle x, d^{(k)}\kappa^\theta \right\rangle_{L_1(\Omega)}, \tag{I.28}$$

il faut qu'il y ait un sens à chercher à extraire de l'information de la dérivée de la fonction de données. En effet, comme le montre l'expression (I.28), la dérivée de l'information n'est autre que l'extraction d'information à partir de la dérivée de la fonction de données.

Certaines applications d'extraction sommative d'informations nécessitent l'utilisation de noyaux dérivables pour obtenir la dérivée de l'information. L'estimation de la dérivée d'un signal image pour extraire les contours des objets projetés sur l'image en est un exemple. On pourrait résumer la situation comme suit : si l'on veut faire une extraction sommative de la dérivée k-ième de l'information, il faut utiliser un noyau sommatif à support borné k fois dérivable.

Notons $\mathbb{S}^k(\Omega)$ l'ensemble des noyaux sommatifs à support borné k fois dérivables ($k \in \mathbb{N}$). On peut remarquer trivialement que $\forall k \geq 1$, $\mathbb{S}^k(\Omega) \subset \mathbb{S}_c(\Omega)$.

L'analogie entre l'extraction sommative d'information et les distributions de Schwartz est directe. Elle se réfère à la théorie des noyaux de Schwartz [56]. Cette théorie expose des résultats sur un cas particulier de distributions qui sont des formes bilinéaires sur l'espace des fonctions test $\mathcal{D}(\Omega)$. De telles distributions peuvent ainsi s'écrire comme des extractions sommatives d'informations (I.12) sous la forme d'opérateurs intégrales [79].

I.3 L'extraction sommative d'informations en traitement du signal numérique

À son origine, le traitement du signal a été développé en vue d'analyser ou modifier des signaux analogiques ; le traitement des signaux numériques était alors présenté comme un cas particulier à fort potentiel.

Depuis la révolution numérique des années 90, il reste peu de processus analogiques en traitement du signal, car la majorité des signaux sont sauvegardés ou transmis sous une forme numérique (et donc échantillonnée). L'exemple de la photographie est symptomatique de ce courant : il reste très peu de photographes utilisant des appareils analogiques, et même ceux-là numérisent ensuite leurs épreuves en vue d'un traitement ultérieur. L'apparition et la popularisation de logiciels tels que Photoshop© a permis au grand public comme aux professionnels de réaliser des opérations de retouche de photos jusque là réservées à des techniciens très expérimentés. La photographie numérique n'est pas le seul

exemple de cette expansion du traitement numérique du signal. L'apparition des CD en 1982 a posé les premières briques du traitement numérique des signaux sonores.

Un signal numérique est un ensemble de valeurs réelles échantillonnées $\{x_1, ..., x_n\}$ provenant de l'observation d'un système physique que l'on souhaite mesurer, contrôler ou analyser. Ces valeurs sont organisées chronologiquement ou spatialement de façon ordonnée si la mesure est systématique ou désordonnée en cas d'observations aléatoires.

S'il existe une grande quantité de problèmes en traitement du signal qui peuvent faire l'objet d'une approche purement numérique, la plupart des problématiques de filtrage, de reconstruction, de déconvolution ou d'analyse sont plus aisées à formaliser dans l'espace continu. Pour cette classe de problèmes, les noyaux sommatifs jouent un rôle primordial de passage d'un problème continu à un problème discret et vice-versa. Un noyau sommatif définit un voisinage pondéré de chaque échantillon numérique, permettant d'étendre la valeur de la mesure aux points voisins : c'est la reconstruction ; ou représente simplement la relation entre le signal original continu et son observation : c'est l'échantillonnage.

Un signal analogique est un ensemble de valeurs réelles $\{x(\omega)|\omega \in \Omega\}$ indexées par l'ensemble référence Ω. En général, Ω est le temps, c'est par exemple le cas pour un signal sonore analogique, mais ça peut aussi bien être un compact de $\mathbb{R}^n$. C'est le cas pour une photographie analogique où Ω est un compact de $\mathbb{R}^2$.

I.3.1 Modélisation de la mesure

Les capteurs sont des dispositifs capables de mesurer des données en vue de les communiquer à d'autres organes du système dans lequel ils sont intégrés. La mesure ou l'acquisition d'une donnée, c'est le processus qui permet d'obtenir une valeur représentant une facette de la réalité. La donnée acquise peut être vue comme une abstraction, une simplification ou une photographie de la réalité. Logiquement, l'extraction d'une information à partir de données doit être le processus qui suit l'acquisition de données. Si on suit cette logique, les mesures sont des données et un processus ultérieur sera dédié à en extraire de l'information. Nous remettons en cause cette vision simplificatrice. Pour nous, le fait de mesurer rentre dans le cadre de l'extraction d'information : la réalité que l'on tente de saisir est alors la donnée tandis que la mesure est l'information. Autrement dit, les éléments du problème d'extraction d'informations sont dans le cas de la mesure :

– données (à traiter) : la réalité,
– information (à extraire) : la mesure.

En traitement du signal, c'est en fait le capteur qui est modélisé par un noyau sommatif. Nous reprenons cette modélisation pour la modélisation de la mesure. Voyons la construction de ce modèle.

Idéalement, l'acquisition d'une donnée à partir de la réalité (ici la donnée du problème d'extraction) $x : \Omega \to \mathbb{R}$ par un capteur placé en ω de Ω est modélisée par la convolution en ω de la réalité x avec l'impulsion de Dirac δ. Ainsi, l'acquisition d'une donnée en ω est modélisée par :

$$\check{x}(\omega) = \int_{\Omega} x(v)\delta(\omega - v)dv = x(\omega), \tag{I.29}$$

où $\check{x}(\omega)$ est la mesure en ω.

Cette "fonction" δ n'est qu'une idéalisation intellectuelle d'une fonction dont on souhaiterait qu'elle possède la propriété suivante :

$$\int_{\Omega} \varphi(v)\delta(v)dv = \varphi(0),\ \forall \varphi \in \mathcal{D}(\Omega). \tag{I.30}$$

Une telle fonction δ n'existe pas. On peut éventuellement la matérialiser par :

$$\delta(\omega) = \begin{cases} 0 & \text{si} \quad \omega \neq 0, \\ +\infty & \text{si} \quad \omega = 0. \end{cases} \tag{I.31}$$

Mais cette représentation est fausse au sens où la fonction (I.31) ne vérifie pas l'égalité (I.30).

On note δ^a l'impulsion de Dirac translatée en a, i.e. c'est la distribution qui à une fonction donnée associe la valeur de cette fonction en a, i.e. $\delta^a(u) = \delta(a-u)$.

On peut réécrire ce modèle comme l'intégrale du produit de la fonction de données x avec l'impulsion de Dirac en ω, i.e. avec δ^ω :

$$\check{x}(\omega) = \langle x, \delta^\omega \rangle_{L_1(\Omega)} = \int_{\Omega} x(v)\delta^\omega(v)dv = x(\omega). \tag{I.32}$$

Le modèle d'échantillonnage de Dirac, même s'il est très utilisé pour une mathématisation de l'acquisition, est très loin de refléter la réalité. En effet, si x est considéré comme la réalité et $\check{x}$, sa mesure, on a, en tout point ω de Ω, $x(\omega) = \check{x}(\omega)$. Traduit en d'autres termes, il y a identité complète entre la réalité à mesurer et la mesure. L'essence même de la mesure est toute autre : mesurer fait généralement intervenir des intégrations locales (temporelles ou spatiales). Une façon de modéliser cette réalité consiste alors à associer la mesure à une intégrale locale (ou une moyenne locale) de la réalité x à mesurer. En supposant que Δ, la largeur de l'intervalle d'intégration, est connue, cela revient à modéliser le capteur par un noyau sommatif uniforme u_Δ, défini par $u_\Delta = \frac{1}{\Delta}\mathbb{1}_{[-\Delta/2,\Delta/2]}$. L'acquisition est alors modélisée par la convolution du signal réel x avec le noyau uniforme u_Δ, i.e.

$$\check{x}(\omega) = \int_{\Omega} x(v)u_\Delta(\omega - v)du = \frac{1}{\Delta}\int_{\omega-\Delta/2}^{\omega+\Delta/2} x(v)dv. \tag{I.33}$$

où $\check{x}(\omega)$ est la mesure en ω. On peut réécrire ce modèle comme suit :

$$\check{x}(\omega) = \langle x, u_\Delta^\omega \rangle_{L_1(\Omega)} = \int_{\Omega} x(v)u_\Delta^\omega(v)du. \tag{I.34}$$

Cependant, la largeur de l'intervalle, si elle existe, est généralement mal connue et la modélisation de la mesure par une simple moyenne locale conduit à une mesure $\check{x}$ souvent très éloignée de la réalité. Faire l'hypothèse que l'agrégation du signal réel par le capteur se fait de façon uniforme est trop optimiste.

En général, un capteur est associé à un accumulateur qui emmagasine le réel x de façon pondérée dans un voisinage de ω. Cette association conduit à définir, autour de chaque point de mesure $\omega \in \Omega$, un voisinage pondéré qu'est le noyau sommatif κ. L'acquisition est alors modélisée par le produit de convolution en ω, de la réalité x avec le noyau sommatif κ, i.e.

$$\check{x}(\omega) = \int_{\Omega} x(v)\kappa(\omega - v)dv = (x * \kappa)(\omega), \tag{I.35}$$

où $\check{x}(\omega)$ est la mesure en ω. On peut réécrire ce modèle comme suit :

$$\check{x}(\omega) = \langle x, \kappa^\omega \rangle_{L_1(\Omega)} = \int_\Omega x(v)\kappa^\omega(v)du. \tag{I.36}$$

On appelle généralement ce noyau sommatif κ, la "réponse impulsionnelle du capteur", car c'est la mesure que délivrerait le capteur, si le signal à mesurer était une impulsion de Dirac.

La modélisation de l'acquisition d'une donnée par l'utilisation d'un noyau sommatif est alors un cas particulier d'extraction sommative d'information. En effet, il est immédiat de constater que :

$$\forall \omega \in \Omega,\ y(\omega) = \check{x}(\omega) = \langle x, \kappa^\omega \rangle_{L_1(\Omega)} = \mathbb{E}_{\kappa^\omega}(x). \tag{I.37}$$

Cet exemple est très intéressant car il donne une interprétation probabiliste au noyau sommatif. Si l'on considère une variable aléatoire sur Ω qui représente l'incertain sur la bonne position ω du capteur, la valeur de la mesure issue du capteur sera donc l'espérance de la fonction de données réelles x suivant la loi qui représente l'incertain sur la position du capteur ω, c'est-à-dire suivant κ^ω. Le noyau sommatif κ^ω représente donc l'incertain sur le positionnement d'un hypothétique capteur ponctuel en ω. On modélise le défaut de la donnée par de l'incertain sur sa position d'acquisition.

I.3.2 Traitement du signal par passages Continu/Discret (C/D)

I.3.2.1 Échantillonnage d'un signal

On dispose d'un signal continu x sur Ω, que l'on souhaite échantillonner afin obtenir la valeur d'une mesure de ce signal $\check{x}_k$ en des points particuliers de Ω : $(\omega_k)_{k=1,..,p}$, appelés points d'échantillonnage, en vue de traiter ce signal par ordinateur, ou d'en créer un enregistrement numérique par exemple.

La modélisation de l'échantillonnage d'un signal est similaire à celle de l'acquisition d'un signal (I.35), c'est-à-dire que l'obtention de la valeur du signal en un point d'échantillonnage ω_k se modélise comme l'acquisition du signal x en ω_k via le noyau sommatif κ^k. On a ainsi, pour tout k de $\{1, ..., p\}$,

$$\check{x}_k = \check{x}(\omega_k) = \int_\Omega x(v)\kappa^k(v)dv = \left\langle x, \kappa^k \right\rangle_{L_1(\Omega)}. \tag{I.38}$$

où κ^k représente la façon dont sont agrégées, en ω_k, les valeurs du signal continu x en vue d'obtenir une valeur de la mesure échantillonnée $\check{x}_k$ de ce signal.

La modélisation de l'échantillonnage d'un signal est donc un cas particulier d'extraction sommative d'informations, exactement comme la modélisation de l'acquisition :

$$\forall k \in \{1, ..., p\},\ y_k = \check{x}_k = \left\langle x, \kappa^k \right\rangle_{L_1(\Omega)} = \mathbb{E}_{\kappa^k}(x). \tag{I.39}$$

Cet exemple permet encore une fois de donner une interprétation probabiliste au noyau sommatif utilisé. Dans ce cas, c'est la position d'échantillonnage qui est une variable aléatoire de loi κ^k. La valeur d'échantillonnage $\check{x}_k$ est la valeur espérée par la fonction de donnée x en ω^k quand cette position suit une loi κ^k.

Rappelons que la modélisation d'une acquisition sans incertitude fait appel à la distribution de Dirac qui modélise la réponse impulsionnelle du capteur "parfait". On peut considérer, de la même manière, le cas d'un échantillonnage "parfait", où on n'a aucune incertitude sur la localisation des points d'échantillonnage. C'est encore une fois la distribution de Dirac que l'on va utiliser. Cette échantillonnage, que l'on appelle l'échantillonnage de Dirac, est obtenu par :

$$\forall k \in \{1, ..., p\},\ y_k = \check{x}_k = \langle x, \delta^{\omega_k} \rangle_{L_1(\Omega)} = x_k. \tag{I.40}$$

I.3.2.2 Reconstruction d'un signal

Le processus de reconstruction est à l'inverse de celui de l'échantillonnage. Reconstruire un signal consiste en l'obtention d'un signal continu $\hat{x}$ sur Ω, à partir d'observations discrètes de ce signal $(x_i)_{i=1,..,n}$ en des points $(\omega_i)_{i=1,..,n}$ de Ω. La reconstruction de $\hat{x}$ est effectuée en associant, à tout point ω de Ω, un noyau sommatif κ^ω. Dans ce cas, le rôle du noyau sommatif utilisé est de collecter les données, i.e. les observations discrètes du signal, autour du point de reconstruction ω, et de les agréger en ω. On utilise pour cela la représentation en distribution (au sens de Schwartz) des observations $(x_i)_{i=1,..,n}$. Cette distribution, notée x^* est appelée distribution des observations. Elle est donnée par :

$$x^* = \sum_{i=1}^{n} x_i \delta^{\omega_i}, \tag{I.41}$$

où δ est la distribution de Dirac.

Ainsi, pour tout ω de Ω, on a :

$$\hat{x}(\omega) = \sum_{i=1}^{n} x_i \kappa^\omega(\omega_i) = \int_\Omega x^*(u)\kappa^\omega(u)du = \langle x^*, \kappa^\omega \rangle_{L_1(\Omega)}. \tag{I.42}$$

La reconstruction d'un signal est donc un cas particulier d'extraction sommative d'informations :

$$\forall \omega \in \Omega,\ y(\omega) = \hat{x}(\omega) = \langle x^*, \kappa^\omega \rangle_{L_1(\Omega)} = \mathbb{E}_{\kappa^\omega}(x^*). \tag{I.43}$$

Cet exemple permet, encore une fois, de donner une interprétation probabiliste au noyau sommatif utilisé. Dans ce cas, c'est la position de reconstruction qui est une variable aléatoire de loi κ^ω. La valeur reconstruite du signal $\hat{x}(\omega)$ est la valeur espérée par la fonction de donnée, qu'est la distribution des observations x^* en ω quand cette position suit une loi κ^ω.

I.3.2.3 Le schéma des méthodes basées sur les passages Continu/Discret

Un grande variété de procédures du traitement du signal dépendent des méthodes de reconstruction et d'échantillonnage d'un signal : le filtrage (stochastique, passe-bas,...), les transformations rigides d'images, la dérivation sont des exemples de telles procédures qui dépendent des interactions entre le discret et le continu. Ces interactions permettent l'obtention d'algorithmes directs ou itératifs, linéaires ou non-linéaires ayant une signification dans le domaine continu.

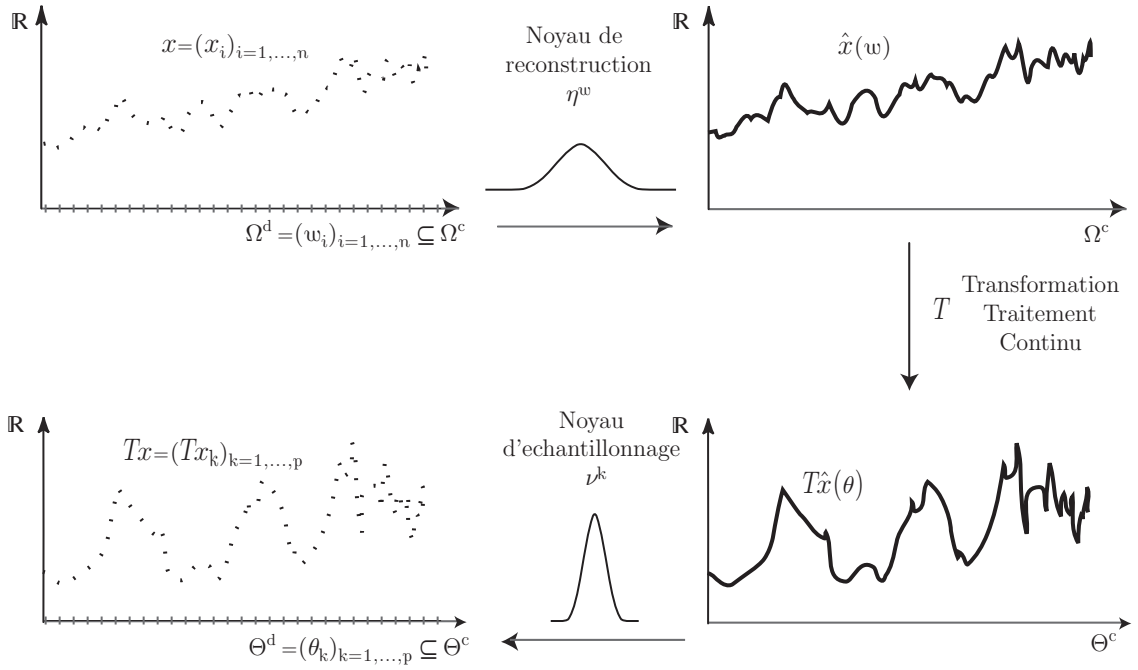


Fig. I.5 : Schéma général du traitement du signal par passages Continu/Discret

Ces méthodes sont toutes basées sur le même schéma : réaliser un algorithme de traitement numérique de signal à partir d'une modélisation analogique du processus de traitement. On part d'un signal numérique x, défini sur $\Omega^d = \{1, ..., n\}$. L'exposant d apposé à Ω signifie que l'on est en discret où chaque indice $i \in \Omega^d$ peut être identifié à un élément ω_i d'un ensemble de référence continu, noté Ω^c (l'exposant c est pour continu). Le signal $x = (x_i)_{i=1,...,n} = (x(\omega_i))_{i=1,...,n}$ est alors reconstruit sur Ω^c, avec les noyaux sommatifs de reconstruction $(\eta^\omega)_{\omega\in\Omega^c}$, par la relation (I.42) en un signal $\hat{x}$, défini pour tout $\omega \in \Omega^c$, par :

$$\hat{x}(\omega) = \sum_{i=1}^{n} x_i \eta^\omega(\omega_i). \tag{I.44}$$

C'est la première étape de la figure I.5.

Le signal reconstruit $\hat{x}$ est ensuite traité, transformé par une opération définie dans l'espace analogique T, qui au signal analogique $\hat{x}$ défini sur Ω^c va retourner un signal analogique $T\hat{x}$ défini sur Θ^c. C'est la deuxième étape de la figure I.5.

Le signal analogique traité, défini par $T\hat{x}(\theta)$, $\forall\theta \in \Theta^c$, est échantillonné avec les noyaux d'échantillonnage $(\nu^k)_{k\in\{1,...,p\}}$. À un noyau d'échantillonnage correspond une position θ_k de l'espace d'échantillonnage Θ^d. Le signal numérique obtenu après reconstruction, traitement et échantillonnage, noté Tx_k pour tout $k \in \{1, ..., p\}$ est obtenu par la relation (I.38) :

$$Tx_k = \int_{\Theta^c} T\hat{x}(\theta)\nu^k(\theta)d\theta. \tag{I.45}$$

C'est la troisième étape de la figure I.5.

Supposons que le traitement T soit une extraction sommative d'informations pour laquelle la fonction de données est $\hat{x}$, la fonction d'informations extraite est $T\hat{x}$ et les noyaux sommatifs utilisés sont $(\kappa^\theta)_{\theta\in\Theta}$; c'est-à-dire $\forall\theta \in \Theta$, $T\hat{x}(\theta) = \int_{\Omega^c} \hat{x}(\omega)\kappa^\theta(\omega)d\omega$,

alors le traitement de $x = (x_i)_{i=1,...,n}$ par passage continu discret et extraction sommative d'informations avec κ^θ est un cas particulier d'extraction sommative d'informations séquentielle à trois noyaux : η^ω, κ^θ et ν^k. Le passage de $x = (x_i)_{i=1,...,n}$ à $T\hat{x} = (T\hat{x}_k)_{k=1,...,p}$ est donc équivalent à une extraction sommative d'information discrète.

I.3.2.4 Adéquation parfaite

Les représentations continues et discrètes d'un même signal doivent être cohérentes entre elles et le passage d'une représentation à l'autre doit se faire avec le souci de respecter cette adéquation. Ces passages sont réalisés par les opérations d'échantillonnage et de reconstruction sur lesquelles il est donc nécessaire de poser des conditions pour avoir cette cohérence. Il n'est pas question de trouver des conditions simultanées sur les noyaux d'échantillonnage et de reconstruction qui permettent d'avoir, après des passages successifs entre le continu et le discret, des représentations qui soient vraies par rapport au signal réel sous-jacent. Une telle adéquation est impossible à réaliser car, par essence, l'échantillonnage fait perdre de l'information de façon irréversible. On peut cependant, quand on connaît les noyaux qui ont échantillonné un signal d'entrée continu quelconque, déterminer des conditions sur les noyaux d'interpolation qui feront que la reconstruction de ce signal sera identique au signal d'entrée. Inversement, si l'opération première est la reconstruction, on peut, quand on connaît les noyaux qui ont interpolé un signal d'entrée discret quelconque, déterminer des conditions sur les noyaux d'échantillonnage qui feront que l'échantillonnage de ce signal sera identique au signal d'entrée aux points d'échantillonnage.

Cette cohérence ou adéquation parfaite est appelé "perfect fit" par Unser [102, 103, 104].

Pour clarifier cela, voyons une version simplifiée du schéma de la figure I.5, où l'opération T est l'identité.

Supposons que l'on connaisse les noyaux de reconstruction η^ω, pour tout $\omega \in \Omega$. Quelles sont les conditions sur les noyaux d'échantillonnage ν^k, pour tout $k \in \Omega^d = \{1, ..., n\}$, pour qu'on ait une cohérence entre les représentations $(x_i)_{i=1,...,n}$, $(\hat{x}(\omega))_{\omega\in\Omega^c}$ et $(\check{x}_i)_{i=1,...,n}$? La cohérence est traduite ici par l'équation : $\forall i \in \Omega^d = \{1, ..., n\}$, $x_i = \check{x}_i$, qui fait intervenir de façon implicite $(\hat{x}(\omega))_{\omega\in\Omega^c}$, par l'intermédiaire de $(\check{x}_i)_{i=1,...,n}$.

Pour déterminer ces conditions, remarquons d'abord, d'après l'expression (I.44), que

$$\forall \omega \in \Omega^c,\ \hat{x}(\omega) = \sum_{i=1}^{n} x_i \eta_i^\omega.$$

Ensuite, d'après l'expression (I.45),

$$\forall k \in \Omega^d = \{1, ..., n\},\ \check{x}_k = \int_{\Omega^c} \hat{x}(\omega)\nu^k(\omega)d\omega.$$

D'où, directement

$$\forall k \in \Omega^d = \{1, ..., n\},\ \check{x}_k = \sum_{i=1}^{n} x_i \int_{\Omega^c} \eta_i^\omega \nu^k(\omega)d\omega.$$

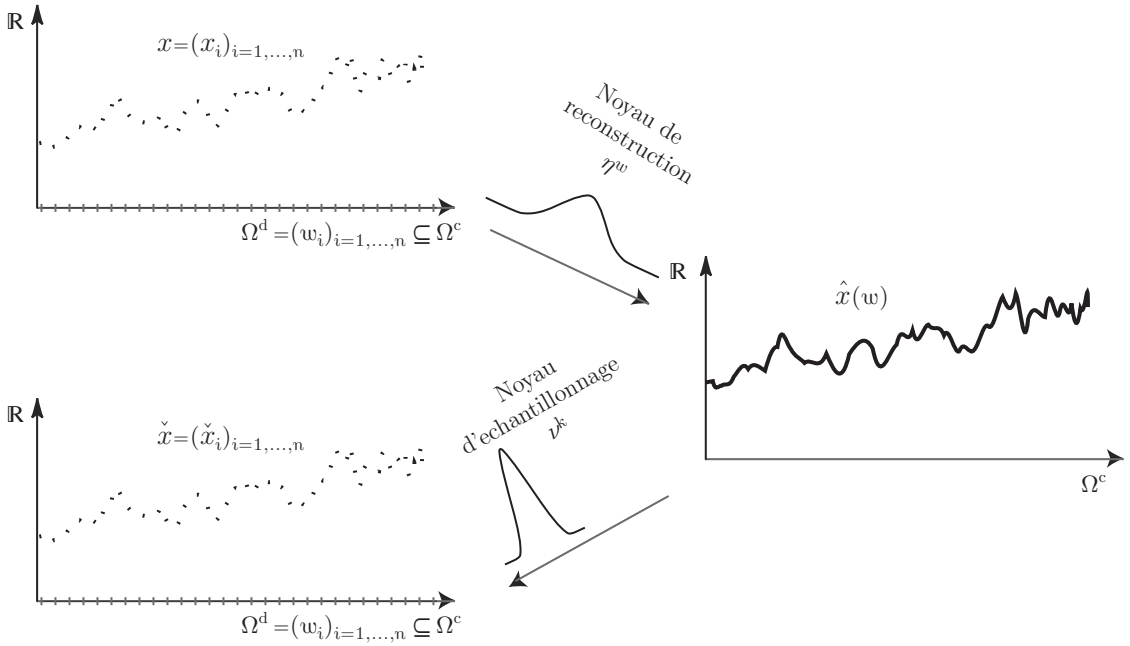


Fig. I.6 : Adéquation parfaite discrète

La condition sur les noyaux d'échantillonnage ν^k, pour tout $k \in \Omega^d = \{1, ..., n\}$, quand on connaît les noyaux de reconstruction $(\eta^\omega)_{\omega\in\Omega^c}$, qui permet d'avoir l'adéquation parfaite entre les représentations discrètes et continues, d'un signal, à l'origine discret, est donc :

$$\int_{\Omega^c} \eta_i^\omega \nu^k(\omega) d\omega = \delta_{ik}, \tag{I.46}$$

où δ_{ik}, qui est le symbole de Kronecker, est défini par :

$$\delta_{ik} = \begin{cases} 1 & \text{si } i = k, \\ 0 & \text{si } i \neq k. \end{cases}$$

Malgré les similitudes de notations, le symbole de Kronecker et l'impulsion de Dirac sont différents.

On pourra dire que les familles des noyaux sommatifs η^ω, pour tout $\omega \in \Omega$ et ν^k, pour tout $k \in \Omega^d = \{1, ..., n\}$ sont discrètement orthogonales lorsqu'ils respectent la condition d'adéquation parfaite discrète (I.46). Le choix de qualifier ces familles de noyaux d'"orthogonales" est lié à l'analogie que l'on peut faire de la condition (I.46) avec la condition d'orthogonalité [27] des polynômes orthogonaux.

Il y a une réelle nécessité à utiliser l'adéquation parfaite discrète pour tous les algorithmes qui font appel aux passages Continu/Discret. Unser a développé, sur cette idée d'adéquation entre les représentations discrètes et continues, des méthodes faisant appel aux noyaux sommatifs B-splines [103, 104]. Ce choix de B-spline est motivé par la grande simplification de la résolution de l'équation (I.46) que ce type de noyau induit. La famille orthogonale à une famille de B-splines est une famille de B-splines dont l'expression est facile à obtenir à partir de la famille des noyaux de reconstruction B-splines choisis. De plus, lorsqu'on paramètre un B-spline, on choisit ses propriétés de continuité sur le noyau et aux différents nœuds d'interpolation. Remarquons que ces bonnes propriétés font des

B-splines des interpolateurs très populaires, très largement utilisés dans les logiciels de CAO industriels.

Une autre forme d'adéquation parfaite peut être proposée, qui cette fois part d'un signal original continu. Si l'adéquation parfaite discrète apparaît très utile dans les applications faisant appel aux passages C/D, on peut s'interroger sur l'utilité d'une adéquation parfaite continue. Définissons la tout de même.

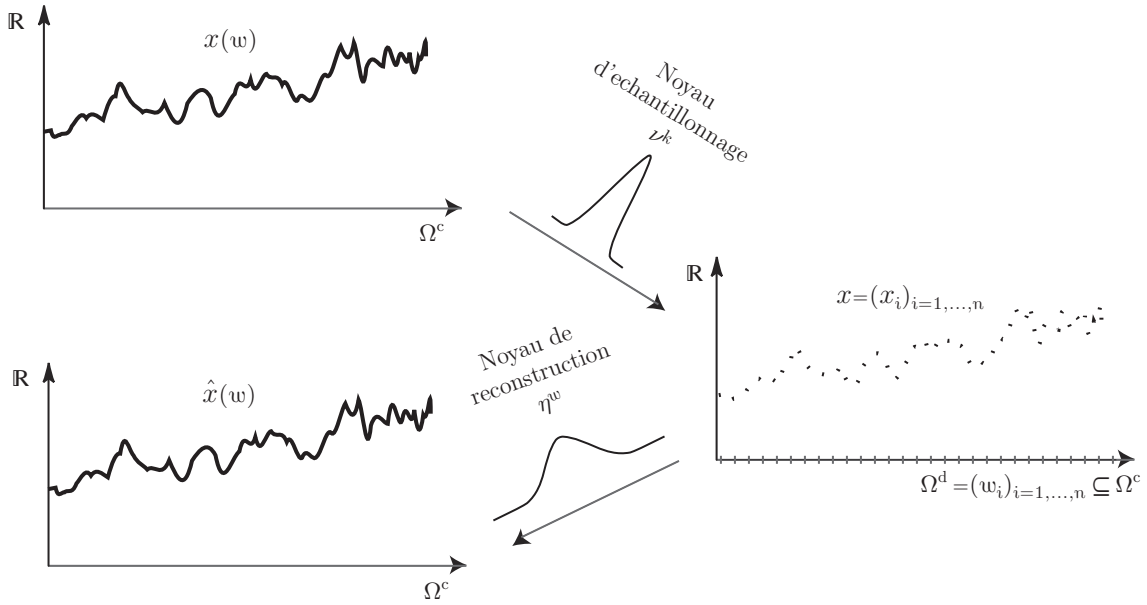


Fig. I.7 : Adéquation parfaite continue

Supposons que l'on connaisse les noyaux d'échantillonnage ν^k, pour tout $k \in \Omega^d = \{1, ..., n\}$. Quelles sont les conditions sur les noyaux de reconstruction η^ω, pour tout $\omega \in \Omega$, pour qu'on ait une cohérence entre les représentations $(x(\omega))_{\omega\in\Omega}$, $(x_i)_{i=1,...,n}$ et $(\hat{x}(\omega))_{\omega\in\Omega}$. La cohérence est traduite ici par l'équation : $\forall\omega \in \Omega$, $x(\omega) = \hat{x}(\omega)$, qui fait intervenir de façon implicite $(x_i)_{i=1,...,n}$, par l'intermédiaire de $(\hat{x}(\omega))_{\omega\in\Omega}$.

Pour déterminer cette condition, remarquons d'abord, d'après l'expression (I.45), que

$$\forall i \in \Omega^d = \{1, ..., n\},\ x_i = \int_{\Omega^c} x(u)\nu^i(u)du.$$

Ensuite, d'après l'expression (I.44),

$$\forall\omega \in \Omega^c,\ \hat{x}(\omega) = \sum_{i=1}^{n} x_i\eta_i^\omega.$$

D'où, directement

$$\forall\omega \in \Omega^c,\ \hat{x}(\omega) = \int_{\Omega^c} x(u) \sum_{i=1}^{n} \nu^i(u)\eta_i^\omega du.$$

La condition sur les noyaux d'échantillonnage η^ω, pour tout $\omega \in \Omega$, quand on connait les noyaux d'échantillonnage ν^k, pour tout $k \in \Omega^d = \{1, ..., n\}$, qui permet d'avoir l'adéquation parfaite entre les représentations discrètes et continues d'un signal, à l'origine

continu, est donc :

$$\sum_{i=1}^{n} \nu^i(u)\eta_i^\omega = \delta^\omega(u), \tag{I.47}$$

où δ^ω est l'impulsion de Dirac en ω.

On pourra dire que les familles des noyaux sommatifs η^ω, pour tout $\omega \in \Omega$ et ν^k, pour tout $k \in \Omega^d = \{1, ..., n\}$ sont continument orthogonales lorsqu'ils respectent la condition d'adéquation parfaite continue (I.47).

I.3.3 Projection tomographique d'un signal image

La modélisation de l'opérateur de projection en tomographie numérique est une application qui rentre dans le cadre présenté à la section précédente I.3.2 des algorithmes de traitement du signal basés sur des passages C/D.

Cette modélisation se base sur la projection en tomographie analogique. Soit $I(u, v)$ une image analogique définie sur un domaine $\Omega^c = U \times V$ de $\mathbb{R}^2$. Sa projection sur un cylindre Θ^c de $\mathbb{R}^+ \times [-\pi, \pi]$, notée $S(\rho, \varpi)$, est une transformation de Radon [49], dont l'expression est :

$$S(\rho, \varpi) = \int_{\mathbb{R}} \int_{\mathbb{R}} I(u, v)\delta(u \cos \varpi + v \sin \varpi - \rho)dudv, \tag{I.48}$$

où δ est la distribution de Dirac. Cette transformation, qui à une image analogique I définie sur Ω^c va associer une image analogique S définie sur Θ^c, est donc un cas particulier des transformations T telles que présentées à la section précédente, où $S = T(I)$.

Les tomographes délivrent des mesures échantillonnées (les détecteurs de radiations sont en quantité fini). Cette transformation (I.48) définie en continu doit donc trouver sa version discrète. On passe pour cela par les méthodes de reconstruction et d'échantillonnage sommatives. La projection discrète ainsi obtenue s'appelle naturellement la transformée de Radon discrète.

Dans la transformation de Radon continue (I.48), le signal d'entrée est une image continue $I(u, v)$. Dans la transformation de Radon discrète, le signal d'entrée est une image numérique $(I_i)_{i=1,...,n}$. À chaque indice $i \in \{1, ..., n\}$ est associé un point ω_i, qui est un couple de points (u_i, v_i) de $\Omega^d \subseteq \Omega^c = U \times V$. L'image reconstruite I en tout $\omega = (u, v)$ de Ω^c est obtenue via le noyau sommatif $\eta^\omega = \eta^{(u,v)}$ par :

$$I(u, v) = \sum_{i=1}^{n} I_i \eta^{(u,v)}(u_i, v_i) = \sum_{i=1}^{n} I_i \eta^\omega(\omega_i). \tag{I.49}$$

On reconnait ici l'expression (I.44). L'expression (I.48) devient :

$$S(\rho, \varpi) = \sum_{i=1}^{n} I_i \int_{\mathbb{R}} \int_{\mathbb{R}} \eta^{(u,v)}(u_i, v_i)\delta(u \cos \varpi + v \sin \varpi - \rho)dudv \tag{I.50}$$

Ce résultat continu de la transformation de Radon à partir d'une image d'entrée numérique $(I_i)_{i=1,...,n}$ est ensuite discrétisée par le noyau sommatif d'échantillonnage ν^k, donc :

$$S_k = \int_{-\pi}^{\pi} \int_{\mathbb{R}^+} S(\rho, \varpi)\nu^k(\rho, \varpi)d\rho d\varpi = \int_{\Theta^c} S(\theta)\nu^k(\theta)d\theta, \tag{I.51}$$

À chaque indice $k \in \{1, ..., p\}$ est associé un point θ_k, qui est un couple de points (ρ_k, ϖ_k) du cylindre de projection Ω. On reconnait l'expression (I.45) avec $\theta = (\rho, \varpi)$. L'expression (I.48) devient :

$$S_k = \sum_{i=1}^{n} I_i r_i^k, \tag{I.52}$$

avec

$$r_i^k = \int_{-\pi}^{\pi} \int_{\mathbb{R}^+} \int_{\mathbb{R}} \int_{\mathbb{R}} \nu^k(\rho, \varpi) \eta^{(u,v)}(u_i, v_i) \delta(u \cos \varpi + v \sin \varpi - \rho) du dv d\rho d\varpi. \tag{I.53}$$

L'opérateur de projection (I.52), qu'est la transformée de Radon discrète du vecteur $I = (I_i)_{i=1,...,n}$ en vecteur $S = (S_k)_{k=1,...,p}$, peut s'exprimer comme un opérateur linéaire avec la matrice de projection à n lignes et p colonnes $R = (r_i^k)_{i=1,...,n;k=1,...,p}$. Cette matrice est appelée matrice de Radon.

I.3.4 Transformation rigide d'une image

Dans beaucoup d'applications, il est nécessaire de recaler deux images en vue de leur comparaison ou de leur fusion. Si les deux images à recaler sont issues de deux capteurs coplanaires, le recalage peut consister en une simple transformation rigide, c'est-à-dire simplement composée d'une translation et d'une rotation. Dans un cas plus général, cette transformation peut être composée avec une homothétie, une distorsion ou une anamorphose. Nous ne considérons ici que la transformation rigide, mais dans tous les cas, le principe est le même.

La transformation rigide d'une image numérique suit le schéma classique des méthodes basées sur le passage continu discret (cf. figure I.3.2.3). La transformation que l'on cherche à effectuer se fait d'une image $(I_i)_{i=1,...,n}$, définie sur un ensemble de pixels $(\omega_i)_{i=1,...,n}$ de Ω vers une image $(I'_k)_{k=1,...,n}$ définie sur un ensemble de pixels $(\omega_k)_{k=1,...,n}$ de Ω. Cette transformation se base sur la transformation continue $t : \Omega \to \Omega$ qui va permuter les éléments de Ω pour transformer la représentation continue $\hat{I}$ sur Ω en une représentation continue transformée $\hat{I}'$ sur Ω.

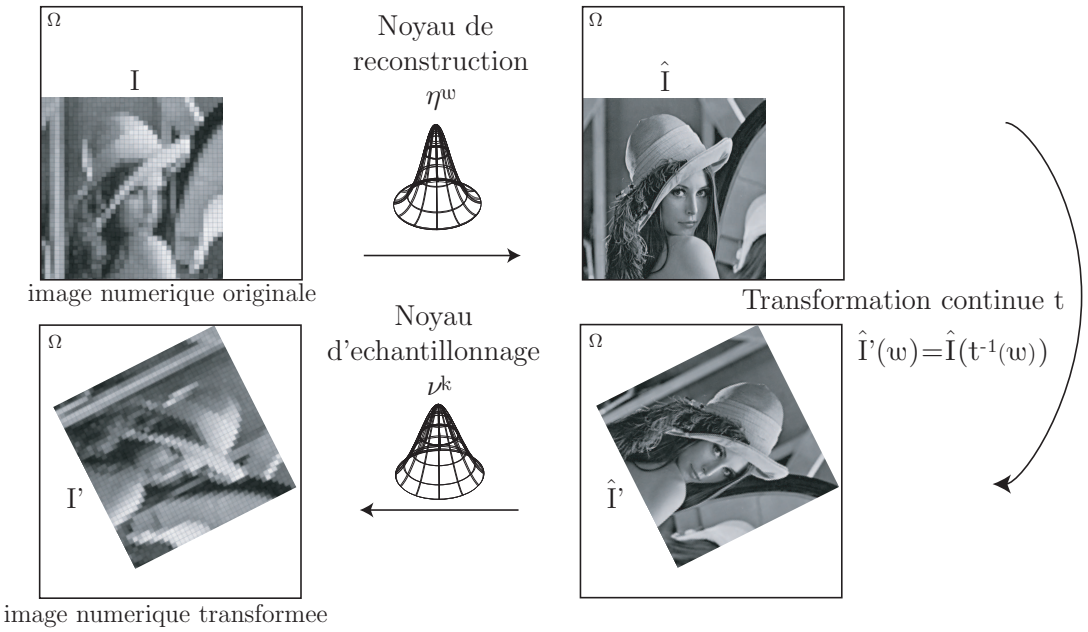


Fig. I.8 : Schéma d'une transformation rigide

Le principe de base des méthodes de transformation d'image est le suivant : pour calculer la valeur en ω_k de l'image transformée I'_k, on essaie d'obtenir la valeur en $t^{-1}(\omega_k)$ de l'image originale $(I_i)_{i=1,\dots,n}$. On passe pour cela à une étape de reconstruction de l'image originale, on obtient $\hat{I}(\omega)$ sur Ω par

$$\hat{I}(\omega) = \sum_{i=1}^{n} I_i \eta^{\omega}(\omega_i). \tag{I.54}$$

La valeur en ω_k de l'image transformée I'_k est donc donnée par

$$I'_k = \hat{I}(t^{-1}(\omega_k)) = \sum_{i=1}^{n} I_i \eta^{t^{-1}(\omega_k)}(\omega_i). \tag{I.55}$$

La plupart des approches s'arrêtent là et font appel à des noyaux d'interpolation ou de reconstruction $\eta^{t^{-1}(\omega_k)}$ aux points inversement transformés $t^{-1}(\omega_k)$ qui peuvent être obtenus, par exemple, par la méthode du plus proche voisin, où la valeur I'_k est simplement la valeur du pixel le plus proche de l'image originale. Ils peuvent être des noyaux bilinéaires produisant une moyenne pondérée (par l'inverse de la distance du pixel au point $t^{-1}(\omega_k)$) des valeurs des 4 (un carré 2×2) pixels les plus proches, où des noyaux bicubiques produisant une moyenne pondérée des valeurs des 16 (un carré 4×4) pixels les plus proches.

La transformation inverse t^{-1} des pixels $(\omega^k)_{k=1,\dots,n}$ de l'image transformée dépend de l'échantillonnage de l'image transformée. Ici c'est un échantillonnage de Dirac (I.40) des pixels $(\omega^k)_{k=1,\dots,n}$ et nous proposons de généraliser cette approche à un échantillonnage par noyaux sommatifs $(\nu^k)_{k=1,\dots,n}$. La transformation inverse t^{-1} qui transforme un des pixels $(\omega_k)_{k=1,\dots,n}$ en un élément de Ω devient une transformation qui transforme un des noyaux sommatifs $(\nu^k)_{k=1,\dots,n}$ en un autre noyau sommatif[2] noté $t^{-1}\nu^k$. L'expression (I.55)

[2]un noyau sommatif soumis à une transformation rigide, i.e. à des rotations et des translations, reste un noyau sommatif

devient alors :

$$I'_k = \sum_{i=1}^{n} I_i \int_\Omega \eta^\omega(\omega_i) t^{-1} \nu^k(\omega) d\omega. \tag{I.56}$$

On peut donc écrire cette transformation rigide comme un opérateur matriciel :

$$I'_k = \sum_{i=1}^{n} I_i t_i^k, \tag{I.57}$$

où

$$t_i^k = \int_\Omega \eta^\omega(\omega_i) t^{-1} \nu^k(\omega) d\omega. \tag{I.58}$$

Ces poids de reconstruction/échantillonnage (car les deux sont utilisés ici) sont des poids sommatifs. En effet,

$$\begin{aligned} \sum_{i=1}^{n} t_i^k &= \int_\Omega (\sum_{i=1}^{n} \eta^\omega(\omega_i)) t^{-1} \nu^k(\omega) d\omega, \\ &= \int_\Omega t^{-1} \nu^k(\omega) d\omega = 1, \end{aligned}$$

car η^ω et $t^{-1}\nu^k$ sont des noyaux sommatifs respectivement discret et continu.

Le poids t_i^k peut être interprété comme la moyenne pondérée du noyau de reconstruction η^ω en ω_i suivant le noyau sommatif d'échantillonnage transformé $t^{-1}\nu^k$.

L'obtention de l'image transformée est une approximation, une agrégation. On a une perte d'information dans la transformation que l'on ne peut pas éviter, car les grilles de pixels des images originale et transformée sont généralement différentes, ce qui nous oblige à reconstruire puis échantillonner. Les reconstruction et échantillonnage sont à l'origine de cette perte d'informations. Les méthodes usuelles sont définies pour minimiser cette perte d'informations lors de la transformation (et aussi pour garder des calculs simples). En particulier, l'involution est préservée, c'est-à-dire que si la transformation est l'identité (donc pas de rotation ni de translation) alors $\forall i \in \{1, ..., n\}$, $I_i = I'_i$. Cela revient donc à choisir les noyaux sommatifs ν^k et η^ω, de telle sorte que la matrice carré $t = (t_i^k)_{i,k=1,...,n}$, soit la matrice identité de dimension n. Autrement dit,

$$\int_\Omega \eta^\omega(\omega_i) \nu^k(\omega) d\omega = \delta_{ik},$$

où δ_{ik} qui est le symbole de Kronecker défini par :

$$\delta_{ik} = \begin{cases} 1 & \text{si } i = k, \\ 0 & \text{si } i \neq k. \end{cases}$$

On reconnait ici la condition d'adéquation parfaite discrète (I.46), tout à fait justifiée par le fait que, sans transformation t, on se retrouve dans le cadre du schéma de la figure I.6.

I.3.5 Filtrage linéaire d'un signal

Le filtrage linéaire, ou filtrage par convolution, regroupe un ensemble de techniques de traitement du signal. Un filtre est caractérisé par sa réponse impulsionnelle. Le filtrage linéaire consiste à convoluer le signal à modifier x avec la réponse impulsionnelle du filtre. C'est pourquoi on appelle souvent *noyau de convolution* cette réponse impulsionnelle. Le filtrage en traitement du signal répond principalement à deux finalités :

- Séparer les signaux utiles des signaux indésirables : ce sont les filtres d'affaiblissement. Les filtres de lissage en sont un exemple.
- Modifier la forme des signaux incidents : ce sont les filtres correcteurs. Les filtres de dérivation par exemple.

On distingue souvent filtres analogiques (filtrage d'un signal continu) et filtres numériques (filtrage d'un signal discret quantifié).

Définition I.7 (Filtre analogique) *Soit x un signal défini sur Ω continu. Soit η une fonction intégrable sur Ω. Le résultat $\tilde{x}(\omega)$ du filtrage du signal x par le filtre de réponse impulsionnelle η est donné, pour $\omega \in \Omega$, par le produit de convolution de x par η en ω, i.e. :*

$$\tilde{x}(\omega) = \int_{\Omega} x(v)\eta(\omega - v)dv = (x * \eta)(\omega). \tag{I.59}$$

On peut remarquer que le produit de convolution peut être remplacé par le produit scalaire sur $L_1(\Omega)$ avec le noyau η^ω. On a ainsi $\tilde{x}(\omega) = \langle x, \eta^\omega \rangle_{L_1(\Omega)}$.

Étant donnée l'analogie complète des expressions (I.35) et (I.59), on peut bien sûr considérer un capteur comme un filtre particulier. Là s'arrête l'analogie, car dans un cas la réponse impulsionnelle est choisie (le filtrage) et dans l'autre cas elle est subie (l'acquisition) et souvent difficile à identifier proprement et précisément.

Définition I.8 (Filtre numérique) *Soit x un signal discret défini sur $\Omega = \{1, ..., n\}$. Soit η une suite sommable (rappel : sommable signifie que $\sum_{i\in\mathbb{Z}} \eta_i < \infty$). Le résultat $\tilde{x}_k$ du filtrage du signal x par le filtre de réponse impulsionnelle η est donné, pour $k \in \{1, ..., n\}$, par le produit de convolution discret de x par η en k, i.e. :*

$$\tilde{x}_k = \sum_{i\in\mathbb{Z}} x_i \eta_{k-i} = (x * \eta)_k. \tag{I.60}$$

Comme dans le cas continu, le produit de convolution peut être remplacé par le produit scalaire sur l_1 avec le noyau η^k, défini pour $k \in \{1, ..., n\}$ par $\forall i \in \mathbb{Z}$, $\eta^k_i = \eta_{k-i}$. Ainsi $\tilde{x}_k = \left\langle x, \eta^k \right\rangle_{l_1(\Omega)}$. Dans le cas numérique, la réponse impulsionnelle du filtre est parfois appelée "vecteur de convolution".

Si certains problèmes de filtrage peuvent faire l'objet d'une spécification purement discrète, la grande majorité des filtres numériques sont des traductions discrètes de filtres analogiques ; c'est le cas du filtre de dérivation appliqué à un signal numérique. Voyons comment définir ces filtres numériques particuliers.

Le but de ce tour de passe-passe est donc le suivant : appliquer un filtre analogique η à un signal numérique $x = \{x_1, ..., x_n\}$. On passe, pour cela, par la représentation en distribution des observations x^* du signal discret x, donnée par l'expression (I.41).

On est toujours dans le cas où les points d'échantillonnage du signal numérique, $(\omega_i)_{i=1,...,n}$, de l'espace continu Ω, sont identifiés à leurs indices $\{1, ..., n\}$. On remarque

alors que le filtrage analogique de x^* par le filtre η^* en un point ω_k est identique au filtrage du signal numérique du signal x par le noyau de convolution numérique défini par $\eta_i = \eta^*(\omega_i)$, pour tout $i \in \{1, ..., n\}$ et $\eta_i = 0$, pour tout $i \in \mathbb{Z}\backslash\{1, ..., n\}$. c'est-à-dire,

$$(x^* * \eta^*)(\omega_k) = (x * \eta)_k. \tag{I.61}$$

En effet,

$$\begin{aligned}
(x^* * \eta^*)(\omega_k) &= \int_\Omega \sum_{i=1}^{n} x_i \delta^i(u) \eta^*(\omega_k - u) du, \\
&= \sum_{i=1}^{n} x_i \int_\Omega \delta^i(u) \eta^*(\omega_k - u) du, \\
&= \sum_{i=1}^{n} x_i \eta^*(\omega_k - \omega_i), \\
&= \sum_{i=1}^{n} x_i \eta_{k-i}, \\
&= (x * \eta)_k.
\end{aligned}$$

Parmi ces filtres (analogiques ou numériques), on pourra distinguer ceux dont la réponse impulsionnelle est un noyau sommatif. Parmi les finalités possibles de ces filtres, certains permettent par exemple un lissage du signal entrant comme les filtres moyenneurs uniformes, les filtres pyramidaux ou les filtres gaussiens. Ceux-ci sont des filtres passe-bas qui coupent plus ou moins les plus hautes fréquences. Ils sont utilisés pour atténuer les bruits d'origines les plus diverses qui polluent le signal. On parle souvent pour ces prétraitements, de débruitage d'un signal.

Tous les noyaux de convolution n'ont pas cette propriété de sommativité. Les filtres utilisés en traitement d'image pour la détection de contour, ont une réponse impulsionnelle qui somme à 0. On pourra citer le filtre de Shen-Castan [95] ou de Canny-Deriche [22].

La seule condition imposée à un noyau de convolution η, c'est son intégrabilité (ou sa sommabilité dans le cas de filtre discret) et non sa sommativité (I.6) ou (I.7). Pour placer le filtrage dans le cadre de l'extraction sommative d'informations il suffit d'écrire tout noyau de convolution comme une combinaison linéaire de noyaux sommatifs.

Théorème I.9. *Tout noyau de convolution peut s'écrire comme la combinaison linéaire de deux noyaux sommatifs.*

Preuve : Soit η un noyau de convolution. On pose $\forall \omega \in \Omega$, $\eta^+(\omega) = \max(0, \eta(\omega))$ et $\eta^-(\omega) = \max(0, -\eta(\omega))$, d'où $\eta = \eta^+ - \eta^-$. Posons $a^+ = \int_\Omega \eta^+(\omega) d\omega$, $a^- = \int_\Omega \eta^-(\omega) d\omega$, $\kappa^+ = \frac{\eta^+}{a^+}$ et $\kappa^- = \frac{\eta^-}{a^-}$. Par construction, κ^+ et κ^- sont sommatifs. η est la combinaison de ces deux noyaux, i.e. $\eta = a^+\kappa^+ - a^-\kappa^-$.

□

La démonstration, pour le cas du filtrage discret, est similaire en remplaçant l'intégrale par une somme.

Cette décomposition d'un noyau de convolution en deux noyaux sommatifs peut s'étendre au filtrage. Autrement dit, le filtrage avec un noyau de convolution quelconque (i.e. intégrable ou sommable) peut être décomposé en combinaison linéaire de filtrages avec noyaux de convolutions sommatifs.

L'intégrale de Lebesgue étant un opérateur linéaire, il en résulte :

$$\int_\Omega x(v)\eta(\omega - v)dv = a^+ \int_\Omega x(v)\kappa^+(\omega - v)dv - a^- \int_\Omega x(v)\kappa^-(\omega - v)dv. \tag{I.62}$$

Donc un filtrage par un noyau de convolution est bien une combinaison linéaire de deux filtrages par noyaux de convolution sommatifs :

$$(x * \eta)(\omega) = a^+(x * \kappa^+)(\omega) - a^-(x * \kappa^-)(\omega). \tag{I.63}$$

La même remarque peut être faite pour le filtrage discret :

$$(x * \eta)_k = a^+(x * \kappa^+)_k - a^-(x * \kappa^-)_k \tag{I.64}$$

On peut donc voir le filtrage continu d'un signal (I.59) comme un cas particulier d'extraction sommative d'informations :

$$\forall \omega \in \Omega,\ \tilde{x}(\omega) = a^+ y^+(\omega) - a^- y^-(\omega) = a^+ \mathbb{E}_{\kappa^{\omega+}}(x) - a^- \mathbb{E}_{\kappa^{\omega-}}(x), \tag{I.65}$$

où $\kappa^{\omega+}$ et $\kappa^{\omega-}$ sont les noyaux sommatifs obtenus à partir du noyau de convolution η^ω et qui forment sa décomposition linéaire.

De même pour le filtrage discret d'un signal (I.60) :

$$\forall k \in \{1, ..., p\},\ \tilde{x}_k = a^+ y_k^+ - a^- y_k^- = a^+ \mathbb{E}_{\kappa^{k+}}(x) - a^- \mathbb{E}_{\kappa^{k-}}(x), \tag{I.66}$$

où κ^{k+} et κ^{k-} sont les noyaux sommatifs obtenus à partir du noyau de convolution η^k et qui forment sa décomposition linéaire.

On peut donc conclure que tout filtrage linéaire de signal peut s'écrire comme une combinaison linéaire d'extractions sommatives d'informations. Cela est dû au fait qu'on peut toujours décomposer un filtre, avec d'une part, la partie positive de la réponse impulsionnelle η du filtre et d'autre part, la partie négative de la réponse impulsionnelle.

Le cas particulier des filtres linéaires à réponse impulsionnelle positive est alors très intéressant, car il s'écrit, à un coefficient a près, comme une extraction sommative d'informations.

Pour le filtrage positif continu d'un signal analogique (I.65) :

$$\forall \omega \in \Omega,\ \tilde{x}(\omega) = a^\omega \mathbb{E}_{\kappa^\omega}(x), \tag{I.67}$$

où κ^ω est un noyau sommatif, tel que $\eta^\omega = a^\omega \kappa^\omega$ et $a^\omega = \int_\Omega \eta^\omega(u)du$.

Pour le filtrage positif discret d'un signal numérique (I.66) :

$$\forall k \in \{1, ..., p\},\ \tilde{x}_k = a^k \mathbb{E}_{\kappa^k}(x), \tag{I.68}$$

où κ^k est un noyau sommatif, tel que $\eta^k = a^k \kappa^k$ et $a^k = \sum_{i=1}^n \eta_i^k$.

I.3.6 Dérivation d'un signal numérique

La dérivation d'un signal numérique $x = \{x_1, ..., x_n\}$ consiste à estimer les valeurs de la dérivée aux points $k = 1, ..., n$. Mais la dérivation est une opération continue. En effet, une règle primaire de l'analyse c'est qu'une fonction dérivable est continue. La contraposée

de cette règle est qu'une fonction discrète (non continue) est non dérivable. Comment procéder pour dériver un signal numérique ?

L'obtention de la dérivée d'un signal numérique $x = \{x_1, ..., x_n\}$ se fait par passage à la représentation en distribution x^* de x, définie par l'expression (I.41). Supposons que l'on ait des observations de la dérivée du signal résumées par leur représentation en distribution (I.41) dx^*. La reconstruction de ce signal dérivé échantillonné, résumé par dx^*, utilise le noyau sommatif η^ω dérivable (donc continu). Ce signal dérivé reconstruit est noté $d\tilde{x}$. Il est donné, pour tout $\omega \in \Omega$, par :

$$d\tilde{x}(\omega) = \langle dx^*, \eta^\omega \rangle_{L_1(\Omega)} . \tag{I.69}$$

Par l'utilisation de l'opération de dérivation au sens des distributions de Schwartz, on a pour $\eta^\omega \in \mathbb{S}^1(\Omega)$,

$$d\tilde{x}(\omega) = - \langle x^*, d\eta^\omega \rangle_{L_1(\Omega)} .$$

Par cette opération unique de dérivation d'un signal discret en en signal dérivé continu, on résume deux opérations : la reconstruction de x^* par η^ω (première étape de la figure I.9), puis la dérivation (seconde étape de la figure I.9).

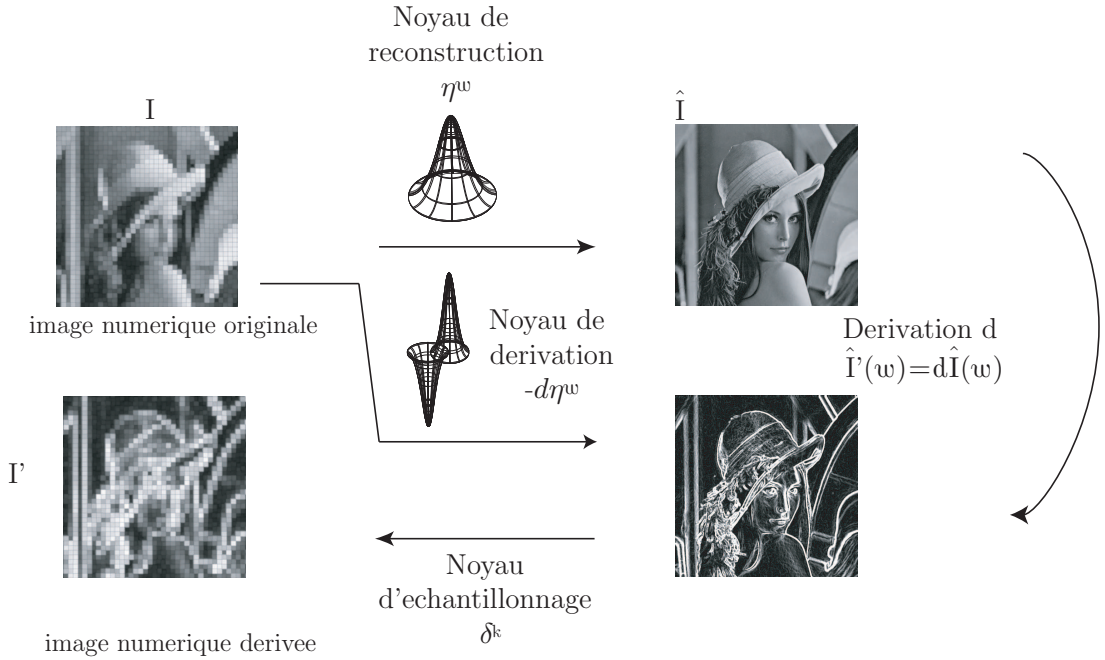


Fig. I.9 : Schéma général de dérivation numérique d'un signal

$d\tilde{x}(\omega)$ peut donc être interprété comme le filtrage de x^* par un filtre à réponse impulsionnelle $-d\eta^\omega$. Un tel filtre est appelé filtre de dérivation. Cette version continue de la dérivation peut être mis dans le cadre de l'extraction sommative d'informations. En effet, pour tout $\omega \in \Omega$,

$$d\tilde{x}(\omega) = a^- \mathbb{E}_{d\eta^{\omega-}}(x) - a^+ \mathbb{E}_{d\eta^{\omega+}}(x). \tag{I.70}$$

où $d\eta^{\omega+}$ et $d\eta^{\omega-}$ sont les noyaux sommatifs obtenus à partir du noyau de convolution $d\eta^\omega$ et qui forment sa décomposition linéaire. Ce noyau de convolution est la réponse impulsionnelle d'un filtre linéaire dérivateur, dont la primitive (i.e. η^ω) est le noyau sommatif de reconstruction qui permet les passages Continu/Discret.

La dérivée numérique du signal x est l'échantillonnage de $d\tilde{x}(\omega)$ obtenu par échantillonnage de Dirac (I.40) $(\delta^k)_{k=1,...,n}$.

$$d\tilde{x}_k = -\left\langle x^*, d\eta^k \right\rangle_{L_1(\Omega)}. \tag{I.71}$$

La dérivation d'un signal numérique est donc un cas particulier d'extraction sommative d'informations (I.66).

$$\forall k \in \{1, ..., n\},\ d\tilde{x}_k = a^- y_k^- - a^+ y_k^+ = a^- \mathbb{E}_{d\eta^{k-}}(x) - a^+ \mathbb{E}_{d\eta^{k+}}(x), \tag{I.72}$$

où $d\eta^{k+}$ et $d\eta^{k-}$ sont les noyaux sommatifs obtenus à partir du noyau de convolution $d\eta^k$ et qui forment sa décomposition linéaire.

L'opération de dérivation peut donc être placée dans le cadre de l'extraction sommative séquentielle d'informations.

I.4 L'extraction sommative d'informations en statistiques

Beaucoup d'outils statistiques s'appuient sur une connaissance plus ou moins exacte de la densité de probabilité sous-jacente à un ensemble d'observation. L'estimation de la densité de probabilité est donc un problème fondamental des statistiques qui a fait l'objet d'une très vaste littérature [97, 89, 96, 80, 101]. Cette estimation s'appuie sur un ensemble fini de n observations $(x_1, ..., x_n)$ de n variables aléatoires $(X_1, ..., X_n)$ iid sur Ω, de loi commune f. C'est cette densité de probabilité f que l'on cherche généralement à estimer ou utiliser indirectement. Le lien entre la densité de probabilité f et la mesure de probabilité P_f des n variables aléatoires est donné par :

$$\forall A \subseteq \Omega, P_f(A) = \int_A f(\omega) d\omega. \tag{I.73}$$

La fonction de répartition F d'une variable aléatoire de mesure de probabilité P_f sur Ω est une représentation équivalente à la densité de probabilité. Cette représentation nécessite qu'il y ait un ordre sur l'ensemble Ω. Prenons par exemple le cas où Ω est l'ensemble des réels. L'ordre sur $\mathbb{R}$ est trivial et la fonction de répartition en ω est donnée par :

$$F(\omega) = P_f(\{u \in \Omega | u \leq \omega\}). \tag{I.74}$$

La représentation de la mesure de probabilité P_f par fonction de répartition et celle par densité de probabilité sont liées par la relation $\forall \omega \in \Omega,\ F(\omega) = \int_{\{u \in \Omega | u \leq \omega\}} f(u) du$. Dans certains cas, il est plus utile, voire plus facile, de connaître la fonction de répartition que la densité de probabilité.

Exemple I.10. *Si l'on cherche à obtenir le quantile d'une variable aléatoire réelle, de fonction de répartition F, il faut inverser cette fonction, car le p-quantile $\omega_{(p)}$, pour $p \in [0, 1]$, est obtenu par la résolution de l'équation $F(\omega_{(p)}) \geq p$.*

Estimer la fonction de répartition à partir d'un ensemble fini de n observations $(x_1, ..., x_n)$ est un problème qui revient fréquemment dans la littérature des statistiques non-paramétriques.

Nous proposons d'étudier les estimateurs de densité de probabilité et de fonctions de répartitions qui rentrent dans le cadre de l'extraction sommative d'informations. Nous nous restreignons dans cette étude au cas où Ω est un sous-ensemble de $\mathbb{R}$.

I.4.1 Estimateur de densité de probabilité de Parzen Rosenblatt

L'estimateur de densité, connu sous le nom d'estimateur à noyau de Parzen Rosenblatt [74, 85], est donné en tout point ω de Ω par :

$$\hat{f}_\kappa(\omega) = \frac{1}{n}\sum_{i=1}^{n}\kappa(\omega - x_i) \tag{I.75}$$

où κ est un noyau sommatif.

Cet estimateur (I.75) s'appuie sur un ensemble de n observations que l'on peut résumer par la distribution empirique [12], définie comme la somme des impulsions de Dirac aux observations, i.e. $e_n = \sum_{i=1}^{n}\delta^{x_i}$.

L'estimateur de Parzen Rosenblatt est un cas particulier d'extraction sommative d'information. En effet, pour tous les ω de Ω, on a :

$$\hat{f}_\kappa(\omega) = y(\omega) = \mathbb{E}_{\kappa^\omega}(x), \tag{I.76}$$

où x la fonction de données est la distribution empirique e_n.

Preuve :

$$\begin{aligned} y(\omega) &= \mathbb{E}_{\kappa^\omega}(e_n), \\ &= \int_\Omega \kappa(\omega - u)e_n(u)du, \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_\Omega \kappa(\omega - u)\delta^{x_i}(u)du, \\ &= \frac{1}{n}\sum_{i=1}^{n}\kappa(\omega - x_i), \\ &= \hat{f}_\kappa(\omega). \end{aligned}$$

□

I.4.2 Estimateur de densité de probabilité par histogramme

L'histogramme est l'estimateur de densité le plus ancien et certainement encore le plus utilisé. D'après [111], il daterait des travaux de John Graunt en 1662.

Le principe de l'histogramme repose sur une représentation de l'ensemble des observations par un nombre fini de classes. Chaque classe est associée à une grandeur, appelée "accumulateur", mesurant sa représentativité dans l'ensemble des observations.

La façon la plus classique de construire les classes consiste à subdiviser le domaine Ω en p intervalles (ou cellules) A_k, de même largeur Δ, et de fonction caractéristique $\mathbb{1}_{A_k}$. Cette largeur est généralement choisie de façon à ce que $\Omega = \cup_{k=1}^{p} A_k$ et $A_i \cap A_j = \emptyset$ si $i \neq j$. Les A_k forment alors une partition exhaustive et exclusive de Ω. Si l'on ordonne les A_k et que l'on appelle ω_k le centre de la cellule A_k, alors $\forall k \in \{1, ..., p\}$, $A_k = [\omega_k - \Delta/2, \omega_k + \Delta/2]$ et $\forall k \in \{1, ..., p-1\}$, $\omega_{k+1} - \omega_k = \Delta$. On appelle Δ le pas de l'histogramme.

Soient $(x_1, ..., x_n)$ les n observations iid de loi commune f. La construction de l'histogramme permettant de représenter f est obtenue en calculant le nombre d'observations

appartenant à chaque cellule A_k. Ce nombre, l'accumulateur Acc_k de la cellule A_k est obtenue par :

$$Acc_k = \sum_{i=1}^{n} \mathbb{1}_{A_k}(x_i). \tag{I.77}$$

Acc_k est un indicateur de représentativité de la cellule A_k dans l'ensemble des n observations au sens où

$$P(A_k) = \frac{Acc_k}{n}.$$

En émettant l'hypothèse que, à l'intérieur d'une cellule A_k, les données se répartissent de façon uniforme, il est possible de bâtir un estimateur empirique de densité par :

$$\hat{f}_\Delta(\omega) = \sum_{k=1}^{p} u_\Delta^k(\omega) P(A_k), \tag{I.78}$$

avec, pour tout $k \in \{1, ..., p\}$,

$$u_\Delta^k(\omega) = \frac{1}{\Delta} \mathbb{1}_{A_k}(\omega), \ \forall \omega \in \Omega.$$

u_Δ^k est une distribution de probabilité, c'est donc un noyau sommatif.

Soit u_Δ^ω, la combinaison convexe des noyaux sommatifs u_Δ^k pondérés par $\mathbb{1}_{A_k}(\omega)$, i.e.

$$\forall u \in \Omega, \ u_\Delta^\omega(v) = \sum_{k=1}^{p} \mathbb{1}_{A_k}(\omega) u_\Delta^k(v). \tag{I.79}$$

On a ainsi que :

$$\hat{f}_\Delta(\omega) = y(\omega) = \mathbb{E}_{u_\Delta^\omega}(x), \tag{I.80}$$

où x, la fonction de données, est la distribution empirique e_n.

Preuve : Voyons d'abord que $\forall k \in \{1, ..., p\}$, $\mathbb{E}_{u_\Delta^k}(e_n) = \frac{Acc_k}{n\Delta}$.

$$\begin{aligned}
\mathbb{E}_{u_\Delta^k}(e_n) &= \int_\Omega u_\Delta^k(u) e_n(u) du, \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_\Omega u_\Delta^k(u) \delta_{x_i}(u) du, \\
&= \frac{1}{n} \sum_{i=1}^{n} u_\Delta^k(\frac{x_i - \omega_k}{\Delta}), \\
&= \frac{1}{n\Delta} \sum_{i=1}^{n} \mathbb{1}_{A_k}(x_i), \\
&= \frac{Acc_k}{n\Delta}.
\end{aligned}$$

Ensuite, l'opérateur d'espérance mathématique $\mathbb{E}$ étant linéaire, on a

$$\mathbb{E}_{u_\Delta^\omega}(e_n) = \sum_{k=1}^{p} \mathbb{1}_{A_k}(\omega) \mathbb{E}_{u_\Delta^k}(e_n).$$

D'où,

$$\begin{aligned}\mathbb{E}_{u_\Delta^\omega}(e_n) &= \sum_{k=1}^{p} \mathbb{1}_{A_k}(\omega)\frac{Acc_k}{n\Delta},\\ &= \sum_{k=1}^{p} u_\Delta^k(\omega)P(A_k),\\ &= \hat{f}_\Delta(\omega).\end{aligned}$$

□

L'estimateur par histogramme est donc un cas particulier de l'estimateur de Parzen Rosenblatt présenté section I.4.1. C'est l'estimateur (I.75) avec le noyau sommatif (I.79).

I.4.3 Estimateur de densité de probabilité par histogramme flou

L'estimateur par histogramme classique (I.78) souffre d'un défaut majeur qu'est l'influence forte du choix de la position et du pas de la partition utilisée pour bâtir l'histogramme. Afin d'atténuer ce défaut, de nombreux auteurs [61, 59, 110] ont proposé de remplacer la partition binaire de l'histogramme classique par une partition floue. Une partition binaire étant un cas particulier de partition floue, les histogrammes flous peuvent être vus comme une généralisation des histogrammes binaires.

Dans cette approche, l'appartenance d'une observation à une cellule A_k ne prend plus valeur dans $\{0,1\}$ mais dans $[0,1]$, réduisant ainsi l'influence de l'arbitraire issu de la définition de la partition de Ω. Pratiquement, remplacer une partition binaire par une partition floue revient à remplacer les fonctions caractéristiques $\mathbb{1}_{A_k}$ de l'expression (I.77) par des fonctions d'appartenance μ_{A_k} à des sous-ensembles flous [116, 30, 28].

Définition I.11. *Un sous-ensemble flou F sur Ω est identifié à sa fonction d'appartenance*

$$\mu_F : \Omega \to [0,1] \tag{I.81}$$

Une partition floue [78] définit donc une partition graduelle de Ω. Cette idée a déjà été utilisé pour réaliser un test du χ^2 [86], pour de l'estimation de mode [99, 100].

Définition I.12. *Soient $\omega_1 < \omega_2 < ... < \omega_p$, les p centres des cellules, tels que $\forall k \neq p$, $m_{k+1} - m_k = \Delta$. Les fonctions d'appartenance aux p cellules $\mu_{A_1}(x)$,$\mu_{A_2}(x)$,...,$\mu_{A_p}(x)$ forment une partition floue de Ω si*

1. $\mu_{A_k}(\omega_k) = 1$,
2. *si* $\omega \notin [\omega_{k-1}, \omega_{k+1}]$, $\mu_{A_k}(\omega) = 0$,
3. $\mu_{A_k}(\omega)$ *est croissant sur* $[\omega_{k-1}, \omega_k]$ *et décroissant sur* $[\omega_k, \omega_{k+1}]$,
4. $\forall \omega \in \Omega$, $\exists k$ *tel que* $\mu_{A_k}(\omega) > 0$,
5. $\forall \omega \in \Omega$, $\sum_{k=1}^{p} \mu_{A_k}(\omega) = 1$,
6. $\forall \omega \in [0, \Delta]$, $\mu_{A_k}(\omega_k - \omega) = \mu_{A_k}(\omega_k + \omega)$,
7. $\forall \omega \in [\omega_k, \omega_{k+1}]$, $\mu_{A_k}(\omega) = \mu_{A_{k-1}}(\omega - \Delta)$ *et* $\mu_{A_{k+1}}(\omega) = \mu_{A_k}(\omega - \Delta)$.

Une propriété intéressante de ce partitionnement flou, c'est qu'il existe un sous-ensemble flou de référence E à partir duquel on peut retrouver les fonctions d'appartenances μ_{A_k} par translation. Autrement dit, toutes les fonctions d'appartenance ont la même forme.

Propriété I.13. *Soit $(\mu_{A_k}(\omega))_{k=1,...,p}$ une partition floue comme présentée dans la définition I.12. Il existe E un sous-ensemble flou, tel que*

$$\forall k \in \{1, ..., p\},\ \mu_{A_k}(\omega) = \mu_E(\omega - \omega_k). \tag{I.82}$$

La propriété I.13 est facilement démontrable avec les points 6 et 7 de la définition d'une partition floue. Un corollaire de cette propriété est que les sous-ensembles flous A_k de cette partition ont tous la même "cardinalité", ou plus exactement la même aire, c'est-à-dire :

Propriété I.14. *Soit $(\mu_{A_k}(\omega))_{k=1,...,p}$ une partition floue comme présentée dans la définition I.12.*

$$\forall k \in \{1, ..., p\},\ \int_\Omega \mu_{A_k}(\omega)d\omega = \Delta. \tag{I.83}$$

Preuve : D'après la propriété I.13, $\forall k \in \{1, ..., p-1\}$, $\int_\Omega \mu_{A_k}(\omega)d\omega = \int_\Omega \mu_{A_{k+1}}(\omega)d\omega = a$. D'où,

$$\sum_{k=1}^{p} \int_\Omega \mu_{A_k}(\omega)d\omega = \sum_{i=1}^{p} a = p.a,$$

Et d'après le point 5 de la définition d'une partition floue, on a :

$$\sum_{k=1}^{p} \int_\Omega \mu_{A_k}(\omega)d\omega = \int_\Omega \sum_{k=1}^{p} \mu_{A_k}(\omega)d\omega = \int_\Omega 1 d\omega = |\Omega|,$$

d'où $a = \frac{|\Omega|}{p}$, avec $|\Omega|$, la mesure de Lebesgue de Ω un sous-ensemble de $\mathbb{R}$, telle que le pas d'échantillonnage Δ vérifie $\Delta = \frac{|\Omega|}{p}$. Donc, $a = \int_\Omega \mu_{A_k}(\omega)d\omega = \Delta$, $\forall k \in \{1, ..., p\}$. □

D'après la propriété I.14, le noyau sommatif κ^k_Δ, défini, $\forall \omega \in \Omega$, par :

$$\kappa^k_\Delta(\omega) = \frac{\mu_{A_k}(\omega)}{\Delta},$$

est sommatif.

Le calcul de l'accumulateur associé à chaque cellule A_k est une simple généralisation de l'expression (I.77) :

$$Acc_k = \sum_{i=1}^{n} \mu_{A_k}(x_i). \tag{I.84}$$

L'estimateur de densité par histogramme flou est défini, par extension de l'estimateur de densité par histogramme binaire, par :

$$\hat{f}_{A_\Delta}(\omega) = \sum_{k=1}^{p} \mu_{A_k}(\omega)\frac{Acc_k}{n\Delta}. \tag{I.85}$$

Afin de replacer cette application d'estimation de densité de probabilité par histogramme flou dans le cadre de l'extraction sommative d'information, on définit, en ω, un

noyau sommatif κ^{ω}_{Δ} comme la combinaison convexe des noyaux sommatifs κ^{k}_{Δ} pondérés par $\mu_{A_k}(\omega)$, i.e.

$$\forall u \in \Omega,\ \kappa^{\omega}_{\Delta}(u) = \sum_{k=1}^{p} \mu_{A_k}(\omega)\kappa^{k}_{\Delta}(u). \tag{I.86}$$

On a ainsi que :

$$\hat{f}_{A_\Delta}(\omega) = y(\omega) = \mathbb{E}_{\kappa^{\omega}_{\Delta}}(x), \tag{I.87}$$

où x la fonction de données est la distribution empirique e_n.
Preuve : Constatons d'abord que $\forall k \in \{1, ..., p\}$, $\mathbb{E}_{\kappa^{k}_{\Delta}}(e_n) = \frac{Acc_k}{n\Delta}$.

$$\begin{aligned}
\mathbb{E}_{\kappa^{k}_{\Delta}}(e_n) &= \int_{\Omega} \kappa^{k}_{\Delta}(u)e_n(u)du, \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_{\Omega} \kappa^{k}_{\Delta}(u)\delta_{x_i}(u)du, \\
&= \frac{1}{n\Delta}\sum_{i=1}^{n} \mu_{A_k}(x_i), \\
&= \frac{Acc_k}{n\Delta}.
\end{aligned}$$

Ensuite, l'opérateur d'espérance mathématique $\mathbb{E}$ étant linéaire, on a

$$\mathbb{E}_{\kappa^{\omega}_{\Delta}}(e_n) = \sum_{k=1}^{p} \mu_{A_k}(\omega)\mathbb{E}_{\kappa^{k}_{\Delta}}(e_n).$$

D'où, directement

$$\mathbb{E}_{\kappa^{\omega}_{\Delta}}(e_n) = \sum_{k=1}^{p} \mu_{A_k}(\omega)\frac{Acc_k}{n\Delta} = \hat{f}_{A_\Delta}(\omega).$$

□

Nous avons étudié les propriétés qualitatives de cet estimateur dans [59] en mettant en avant, dans des expériences illustratrices, les propriétés de plus grande robustesse de cet estimateur au choix de la partition par rapport à l'histogramme classique. Dans l'article [61] nous avons démontré la convergence en MSE de cet estimateur. La borne de convergence MISE, obtenue dans l'article [110] de Waltman et al., a été affinée pour définir la largeur de bande optimale de cet estimateur par la méthode de validation croisée [89, 96]. Il est important de noter que sa définition est très proche de celle présentée dans [90, 43] sous le nom de *binned kernel estimator*.

Souvent, l'estimateur de Parzen Rosenblatt est présenté comme une extension de la méthode des histogrammes utilisant une partition très fine de Ω et dont chaque cellule serait identifiée à son centre. Cette nouvelle écriture montre clairement que l'estimateur par histogramme flou n'est autre qu'un estimateur de Parzen Rosenblatt (I.75) utilisant le noyau sommatif défini par (I.86).

I.4.4 Estimateurs de fonction de répartition

Soit un ensemble fini de n observations $(x_1, ..., x_n)$ de n variables aléatoires iid, de fonction de répartition commune F.

Tout comme il existe un estimateur à noyau de Parzen Rosenblatt de la densité de probabilité f, il existe un estimateur à noyau de Parzen Rosenblatt de la fonction de répartition F, donné en tout point ω de Ω, par :

$$\hat{F}_\kappa(\omega) = \int_{\{u\in\Omega | u\leq\omega\}} \hat{f}_\kappa(u)du = \int_{\{u\in\Omega | u\leq\omega\}} \frac{1}{n}\sum_{i=1}^{n} \kappa(u - x_i)du \tag{I.88}$$

où κ est un noyau sommatif. Cet estimateur est, par définition, l'intégrale de l'estimateur de densité de probabilité de Parzen Rosenblatt.

L'ensemble des n observations peut être résumé par la fonction de répartition empirique [12], définie comme la somme des fonctions échelons aux observations, i.e. $E_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} H^{x_i}(\omega)$. La fonction échelon, également appelée fonction Heaviside, est définie en $a \in \Omega$ par :

$$\forall\omega \in \Omega,\ H^a(\omega) = \begin{cases} 0 & \text{si} \quad \omega < a \\ 1 & \text{si} \quad \omega \geq a. \end{cases} \tag{I.89}$$

C'est une primitive de la distribution δ^a de Dirac en a au sens des distributions, i.e. $\forall\omega \in \Omega$, $H^a(\omega) = \int_{\{u\in\Omega | u\leq\omega\}} \delta^a(u)du$. On peut en déduire que la fonction de répartition empirique E_n est une primitive de la distribution empirique e_n, i.e.

$$E_n(\omega) = \int_{\{u\in\Omega | u\leq\omega\}} e_n(u)du. \tag{I.90}$$

L'estimateur à noyau de fonction de répartition de Parzen Rosenblatt est un cas particulier d'extraction sommative d'information. Ainsi, pour tout ω de Ω, on a :

$$\hat{F}_\kappa(\omega) = y(\omega) = \mathbb{E}_{\kappa^\omega}(x). \tag{I.91}$$

où x, la fonction de données, est la fonction de répartition empirique E_n.

Preuve : En introduisant l'expression (I.76) ($\hat{f}_\kappa(u) = \mathbb{E}_{\kappa^u}(e_n) = \int_\Omega \kappa(u-v)e_n(v)dv$) dans l'égalité de droite des équations (I.91), on a

$$\begin{aligned}
\hat{F}_\kappa(\omega) &= \int_{\{u\in\Omega | u\leq\omega\}} \Big(\int_\Omega \kappa(u - v)e_n(v)dv \Big)du, \\
&= \int_{\{u\in\Omega | u\leq\omega\}} \Big(\int_\Omega \kappa(v)e_n(u - v)dv \Big)du, \\
&= \int_\Omega \Big(\int_{\{u\in\Omega | u\leq\omega\}} e_n(u - v)du \Big)\kappa(v)dv, \\
&= \int_\Omega E_n(\omega - v)\kappa(v)dv, \\
&= \int_\Omega E_n(v)\kappa(\omega - v)dv, \\
&= \mathbb{E}_{\kappa^\omega}(E_n).
\end{aligned}$$

□

Les estimateurs de densité de probabilité par histogramme flou $\hat{f}_{A_\Delta}$ et binaire $\hat{f}_\Delta$ sont des cas particuliers de l'estimateur de densité de probabilité de Parzen Rosenblatt $\hat{f}_\kappa$. On peut donc également leur associer des estimateurs de fonction de répartition par histogramme flou et binaire, définis par :

$$\hat{F}_{A_\Delta}(\omega) = \int_{\{u\in\Omega | u\leq\omega\}} \hat{f}_{A_\Delta}(u)du, \tag{I.92}$$

$$\hat{F}_{\Delta}(\omega) = \int_{\{u\in\Omega | u\leq\omega\}} \hat{f}_{\Delta}(u)du. \tag{I.93}$$

$\hat{F}_{A_\Delta}$ est l'estimateur de fonction de répartition par histogramme flou et $\hat{F}_\Delta$ est l'estimateur de fonction de répartition par histogramme binaire.

Ces estimateurs sont des cas particuliers d'extractions sommatives d'informations. Ainsi, pour tout ω de Ω, on a :

$$\hat{F}_{A_\Delta}(\omega) = y(\omega) = \mathbb{E}_{\kappa^\omega_\Delta}(x), \tag{I.94}$$

$$\hat{F}_{\Delta}(\omega) = y(\omega) = \mathbb{E}_{u^\omega_\Delta}(x), \tag{I.95}$$

où x la fonction de données est la fonction de répartition empirique E_n, κ^ω_Δ est le noyau sommatif défini par (I.86) et κ^ω_Δ est le noyau sommatif défini par (I.79).

I.4.5 Une autre formulation des estimateurs de densité de probabilité

Lorsque nous avons proposé les estimateurs de densité de probabilité (I.76), (I.80) et (I.85) comme des cas particuliers d'extractions sommatives d'informations, nous avons intentionnellement ignoré le fait que la distribution empirique e_n n'est pas une fonction mais une distribution au sens de Schwartz. Or nous n'avons défini l'extraction sommative d'informations que pour des données de type fonctions. Les formulations de ces estimateurs sont tout de même correctes, car la formulation intégrale avec l'impulsion de Dirac (I.30), bien qu'idéalisée, existe.

Deux alternatives s'offrent à nous pour faire rentrer rigoureusement ces estimateurs dans le cadre de l'extraction sommative d'informations. Soit étendre cette approche à des "fonctions" de données qui seraient, non plus seulement des fonctions, mais des distributions identifiables à des formes intégrables (I.26) comme l'est la distribution de Dirac. Soit reformuler l'extraction sommative d'informations appliquée à la distribution empirique e_n.

La première solution est triviale. La seconde, la reformulation, s'appuie sur la dérivation. En effet, si la fonction échelon H est la primitive de l'impulsion de Dirac δ, δ est la dérivée de H au sens des distributions, i.e.

$$\forall\varphi\in\mathcal{D}(\Omega),\ \langle\delta,\varphi\rangle_{L_1(\Omega)} = \langle dH,\varphi\rangle_{L_1(\Omega)} = -\langle H, d\varphi\rangle_{L_1(\Omega)}. \tag{I.96}$$

Ceci est aussi vrai pour la distribution empirique et la fonction de répartition empirique, i.e.

$$\forall\varphi\in\mathcal{D}(\Omega),\ \langle e_n,\varphi\rangle_{L_1(\Omega)} = -\langle E_n, d\varphi\rangle_{L_1(\Omega)}. \tag{I.97}$$

Cette propriété, vraie pour les fonctions test, l'est aussi pour des noyaux sommatifs une fois dérivables et à supports bornés :

$$\forall\kappa^\omega\in\mathbb{S}^1(\Omega),\ \langle e_n,\kappa^\omega\rangle_{L_1(\Omega)} = -\langle E_n, d\kappa^\omega\rangle_{L_1(\Omega)}. \tag{I.98}$$

La reformulation en extraction sommative d'informations n'est pas directe, car la dérivée d'un noyau sommatif n'est pas sommative. L'étape suivante s'inspire du filtrage (théorème I.9, section I.3.5). $d\kappa^{\omega}$ peut s'écrire comme une combinaison linéaire de deux noyaux sommatifs, i.e.

$$\forall \kappa^{\omega} \in \mathbb{S}^1(\Omega),\ d\kappa^{\omega} = a^{+} d\kappa^{\omega+} - a^{-} d\kappa^{\omega-}. \tag{I.99}$$

On a donc

$$\forall \kappa^{\omega} \in \mathbb{S}^1(\Omega),\ \langle e_n, \kappa^{\omega} \rangle_{L_1(\Omega)} = a^{-} \left\langle E_n, d\kappa^{\omega-} \right\rangle_{L_1(\Omega)} - a^{+} \left\langle E_n, d\kappa^{\omega+} \right\rangle_{L_1(\Omega)}, \tag{I.100}$$

où $d\kappa^{\omega+}$ et $d\kappa^{\omega-}$ sont des noyaux sommatifs.

Ainsi, les estimateurs de densité de probabilité de Parzen Rosenblatt, par histogramme binaire ou flou, peuvent s'écrire comme une combinaison linéaire d'extractions sommatives d'informations, car

$$\forall \kappa^{\omega} \in \mathbb{S}^1(\Omega),\ \mathbb{E}_{\kappa^{\omega}}(e_n) = a^{-} \mathbb{E}_{d\kappa^{\omega-}}(E_n) - a^{+} \mathbb{E}_{d\kappa^{\omega+}}(E_n). \tag{I.101}$$

I.5 Indices de comportement d'un noyau sommatif en extraction sommative d'informations

Le comportement d'un noyau sommatif en extraction sommative d'informations dépend, d'une manière évidente, de sa forme. On appréhende intuitivement qu'un noyau sommatif très étalé mais de faible hauteur, aura un comportement, en extraction sommative d'informations, totalement différent d'un noyau sommatif très concentré en un point et qui monte très haut. En effet, l'extraction sommative d'informations étant une somme pondérée par le noyau sommatif de la fonction de donnée, la répartition des poids sur la fonction de donnée a une influence primordiale. Nous conjecturons que c'est la spécificité du noyau sommatif qui caractérise le mieux le comportement d'un noyau sommatif en extraction sommative d'informations. La spécificité d'un noyau sommatif est sa capacité à être concentré sur un ensemble de longueur minimale. La longueur évoquée ici est la mesure de Lebesgue dans le cas continu et le cardinal de l'ensemble dans le cas discret. La distribution de Dirac (I.31) est le noyau sommatif, le plus spécifique, tandis que le noyau sommatif uniforme est le moins spécifique. Une mesure de spécificité d'un noyau sommatif quelconque se doit donc de capturer une idée de distance entre la distribution de Dirac et le noyau sommatif.

Il n'existe pas, dans la littérature, d'indice naturel de spécificité d'un noyau sommatif caractérisant son comportement en extraction sommative d'informations. Cependant, dû à l'interprétation probabiliste d'un noyau sommatif, des indices dits *d'information* des distributions de probabilité peuvent être de bons candidats pour caractériser le comportement d'un noyau sommatif. En 1948, Shannon a proposé une mesure d'information des distributions de probabilité, connu sous le nom d'entropie de Shannon [94].

Définition I.15. *Soit κ une distribution de probabilité sur Ω, l'entropie de Shannon de κ est définie par :*

$$H(\kappa) = -\int_{\Omega} \kappa(\omega) \log(\kappa(\omega)) d\omega,\ \textit{dans le cas continu}, \tag{I.102}$$

$$H(\kappa) = -\sum_{i=1}^{n} \kappa_i \log(\kappa_i),\ \textit{dans le cas discret}. \tag{I.103}$$

avec la convention suivante : $0 \log 0 := 0$.

Cet indice est l'unique solution d'un ensemble d'axiomes [52]. La rigueur dans la construction de cet indice lui a valu sa grande popularité. Rényi [82, 81] a probablement été le premier à considérer des modifications naturelles aux axiomes de Shannon. Ces modifications ont fait naitre de nouveaux indices d'entropie. L'article de Morales et al. [67] est une étude très complète et utile de nombreux indices d'information.

Quelquefois, la variance de la variable aléatoire associée à la distribution de probabilité qu'est le noyau sommatif, qui n'est ni un indice de spécificité ni un indice d'informations, est considéré comme un indice du comportement du noyau sommatif. La variance, qui mesure la distance moyenne entre les observations et la valeur moyenne du résultat de l'expérience, est un indice de dispersion des observations. Prenons un exemple simple pour illustrer l'affirmation que la variance n'est pas un indice de spécificité du noyau sommatif.

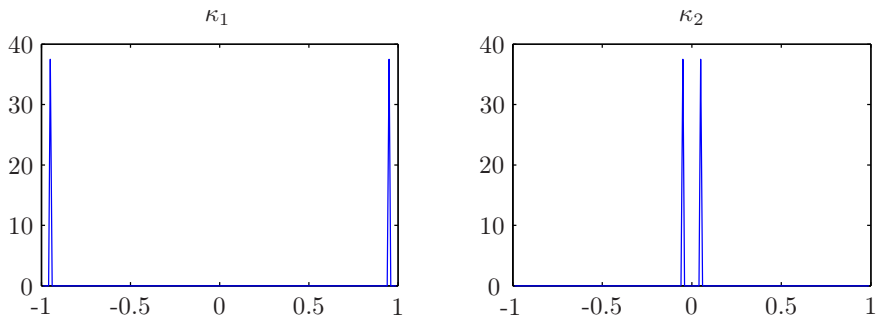


Fig. I.10 : Variance et spécificité

On observe, en figure I.10, deux noyaux sommatifs κ_1 et κ_2. Chacun est une mixture de deux noyaux sommatifs identiques, placés en deux modes différents. Pour κ_1, les modes sont $\{-0.95, 0.95\}$. Pour κ_2, les modes sont $\{-0.05, 0.05\}$. Malgré des positions de modes différents, les ensembles où sont concentrés les poids de κ_1 et κ_2 ont la même longueur. Donc κ_1 et κ_2 ont la même spécificité. Cependant, la variance de κ_1 est plus grande que la variance de κ_2, donc la variance n'est pas un indice de spécificité au sens où nous avons défini la spécificité.

Nous prétendons en fait que la variance ne mesure pas le comportement d'un noyau sommatif mais qu'un indice de spécificité le ferait. L'indice qui se rapproche le plus de nos attentes pour un indice de spécificité d'un noyau sommatif est l'indice de *peakedness* de Birnbaum [10, 35]. Cet indice mesure à quel point une distribution de probabilité est pointue et donc à quel point elle se rapproche de la distribution de Dirac.

Définition I.16. *Soit κ une distribution de probabilité sur Ω, le peakedness de Birnbaum de κ est défini par :*

$$\Gamma(\kappa) = \int_{\Omega} \Big(\int_{\{x \in \Omega | \kappa(x) \geq \kappa(\omega)\}} \kappa(x) dx \Big) d\omega, \quad \textit{dans le cas continu,} \tag{I.104}$$

$$\Gamma(\kappa) = \sum_{i=1}^{n} \sum_{j=i}^{n} \kappa_{(j)}, \quad \textit{dans le cas discret,} \tag{I.105}$$

où $(\cdot)$ est une permutation des indices de sorte que $\kappa_{(1)} \leq \kappa_{(2)} \leq ... \leq \kappa_{(n)}$.

I.6 Défauts et insuffisances de l'extraction sommative d'informations

I.6.1 Choix du noyau sommatif : objectif ou subjectif ?

Dans toutes les méthodes d'extraction sommative d'informations que nous avons présentées, l'étape du choix du noyau sommatif est cruciale. C'est de ce choix que va dépendre le comportement de la méthode. Suivant les écoles ou les domaines d'utilisation, l'utilisateur fera un choix basé sur des critères qui se veulent objectifs, mais dont l'arbitraire ne peut jamais être totalement écarté.

Prenons l'exemple des critères de Canny pour le choix d'un noyau optimal de détection de contour [13, 22]. Des modifications et des discussions de ces critères ont permis de développer des méthodes qui utilisent d'autres noyaux que celui de la méthode de Canny-Deriche, comme dans le cas avec l'utilisation du noyau de Shen-Castan [95]. Le noyau de Canny-Deriche est celui qui respecte le mieux les critères de Canny, celui de Shen-Castan est celui qui respecte le mieux les critères de Shen. Est donc introduit ici une part d'arbitraire et de subjectivité dans le choix des critères.

Certains critères, considérés comme objectifs par les praticiens, sont souvent basés sur des théorèmes mal interprétés. Ainsi, en traitement du signal, le théorème de la limite centrale est à l'origine du choix du noyau gaussien dans beaucoup d'applications. Ce théorème nous dit que la moyenne d'une variable aléatoire réelle quelconque est, elle même, une variable aléatoire de loi gaussienne. Cette propriété remarquable est vraie quelle que soit la loi de la variable aléatoire sous-jacente. Elle ne donne donc aucune information sur cette variable aléatoire. La moyenne d'un lancé de dé suit une loi gaussienne centrée en 3.5, mais le lancé de dé ne suit pas une loi gaussienne ; le lancé de dé suit une loi uniforme. L'acquisition par un voltmètre d'un courant à 30V ne suit pas nécessairement une loi gaussienne alors que sa moyenne suit une loi gaussienne centrée en 30V.

En statistique, certains choix sont présentés comme étant basés sur des critères objectifs. Ainsi, en statistique non-paramétrique, le noyau d'Epanechnikov $Ep(\omega) = \frac{3}{4}(1 - \omega^2)\mathbb{1}_{|\omega|\leq 1}$ est souvent conseillé pour toute application à noyau, que ce soit pour les estimateurs de Parzen Rosenblatt présentés en sections I.4.1 et I.4.4 ou pour de la régression par noyau [97, 68]. Ceci est dû au fait que le noyau d'Epanechnikov minimise l'erreur d'estimation AMISE (Asymptotic Mean Integrated Squared Error) de l'estimateur à noyau de Parzen Rosenblatt $\hat{f}_\kappa$ de densité de probabilité, dont l'expression est

$$AMISE(\hat{f}_\kappa) \approx \frac{1}{4}\sigma_\kappa^4 \int_\Omega (f''(\omega))^2 d\omega + \frac{\int_\Omega \kappa^2(\omega)d\omega}{n} \text{ où } \sigma_\kappa^2 = \int_\Omega \omega^2 \kappa(\omega) d\omega.$$

Mais ce critère AMISE est encore une fois discutable et est souvent discuté, comme le montrent les travaux de Luc Devroye et al. [24, 25] sur des critères basés sur la norme L_1 pour quantifier la distance moyenne entre l'estimateur et la valeur de la densité. Des approches plus récentes, dites combinatoires, ont également été développées par Luc Devroye [26]. On peut aussi citer Kanazawa [50] pour un choix de noyau basé sur la distance de Hellinger. Pour la distance de Kullback-Leibler, on peut se référer au travaux de Rodriguez et Van Ryzin [84]. Kim et Van Ryzin [51] ont étudié une distance basée sur la norme uniforme.

Dans toute méthode d'extraction sommative d'informations, le choix d'un noyau unique apparaît souvent comme une hypothèse trop forte. Relacher cette hypothèse c'est envi-

sager l'utilisation d'une famille de noyaux sommatifs. Les noyaux sommatifs sont des distributions de probabilité. Les théories des probabilités imprécises apparaissent donc comme l'outil adéquat. En effet, un modèle de probabilité imprécise est équivalent à une famille de modèles de probabilités précises.

I.6.2 Les théories des probabilités imprécises

I.6.2.1 Historique de l'imprécis

Il nous semble intéressant de discuter brièvement des interprétations des théories de l'incertain pour comprendre les origines de la confusion historique entre les notions d'incertain et d'imprécis.

La théorie des probabilités classiques, peu importe son interprétation objective ou subjective, est une théorie de l'incertain, précise. Elle quantifie l'incertitude sur un évènement A par une valeur précise $P(A)$ comprise entre 0 et 1.

L'interprétation de la quantité $P(A)$ est au cœur de nombreux, très nombreux débats dont l'utilité est souvent inversement proportionnelle à l'agressivité. Nous discutons ici, de façon dédramatisée, de deux écoles qui se livrent bataille : l'école objectiviste et l'école subjectiviste. La "compétition" entre ces deux écoles s'avère quelques fois constructive avec des chercheurs comme James Berger [7, 16, 8, 2], Frank Coolen [1, 15], Glenn Shafer [93], Peter Walley [108] ou Didier Dubois et Henri Prade [36, 33] qui étudient de façon impartiale les deux interprétations.

L'interprétation objective des probabilités est basée sur des faits. Cela ne fait pas intervenir de croyance quelconque. On s'appuie sur le monde, dont la complexité est infinie, pour évaluer l'incertain. L'évaluation d'une probabilité dépend d'une infinité de paramètres pouvant prendre une infinité de valeurs. Tout modèle d'évaluation de probabilité simplifie cette réalité et ne fait qu'approximer une valeur hypothétiquement vraie de $P(A)$, en rendant finie la complexité infinie du monde qui entoure toute expérience. Prenons l'exemple simple de l'évaluation de la probabilité de tirer le Roi de Pique d'un jeu de 32 cartes. Cette évaluation ne dépend finalement que d'un facteur : le nombre de cartes. L'approximation de la réalité faite par le modèle convient tout à fait à ce problème. Prenons maintenant l'exemple de l'évaluation de la probabilité qu'un malade atteint d'un cancer guérisse. La guérison de ce malade dépend de nombreux facteurs dont beaucoup sont mal mesurés ou ne sont même pas encore identifiés, car la recherche sur le cancer ne les a pas envisagés. L'approximation de la réalité faite pour évaluer cette probabilité ne sera jamais suffisante pour obtenir la vraie valeur de cette probabilité.

L'interprétation subjective des probabilités est basée sur la capacité du cerveau humain à évaluer $P(A)$ à partir de ses croyances, de ses connaissances du monde. On appelle également cette interprétation, les probabilités épistémiques car elles dépendent aussi de notre état de connaissance du monde. L'évaluation subjective d'une probabilité dépend d'une infinité de paramètres, difficilement identifiables et quantifiables. Ces paramètres dépendent du fonctionnement du cerveau humain ou du fonctionnement de notre raisonnement. Peut-être que des tentatives de simplification du fonctionnement du cerveau humain, telle que celle d'Emmanuel Kant dans sa Critique de la Raison Pure (1778), permettront un jour de bien définir ces facteurs, de modéliser parfaitement les rouages de notre raisonnement et donc de notre évaluation de l'incertain.

La simplification de la réalité du monde, nécessaire à l'évaluation objective d'une pro-

babilité, a pu donner l'illusion à des générations de probabilistes (objectivistes) que la théorie des probabilités précises était parfaitement à même de refléter le monde ou l'environnement d'une expérience afin d'en évaluer les incertitudes. Appréhender une probabilité comme une quantification subjective de l'incertain a facilité l'idée que la valeur précise qu'on attribuait à $P(A)$ n'était qu'un choix arbitraire fait dans un ensemble de valeurs potentielles que pourraient prendre cette probabilité. Admettre les faiblesses du cerveau humain (on prend tous les jours conscience de son imperfection), admettre notre propre ignorance, est finalement plus facile que d'admettre les défauts des méthodes objectivistes d'évaluation de l'incertain.

Dans le cadre de l'interprétation objective, certains modèles ont essayé de prendre en compte tant bien que mal l'imprécision. Ce n'est qu'à la suite des travaux de Bruno de Finetti [18, 19] sur une théorie subjective des probabilités au moins aussi rigoureuse que la théorie objective de Andreï Kolmogorov [53] que des théoriciens comme Peter Williams [113] et, plus récemment, Peter Walley [105] ont pu fournir un cadre théorique rigoureux à l'évaluation imprécise de l'incertain.

I.6.2.2 La théorie générale des prévisions cohérentes de Peter Walley

Le livre de Peter Walley [105] sur le modèle des prévisions inférieures et supérieures cohérentes est un exposé complet des théories des probabilités imprécises, accessible et clair. Par *théories des probabilités imprécises*, nous entendons toutes les théories qui évaluent l'incertain par des valeurs non précises, non ponctuelles. Peter Walley considère un modèle très général où un agent est soumis à de l'incertain sur la valeur prise par une variable X dans Ω, où Ω est un ensemble fini. Sa formulation est basée sur l'interprétation subjective des probabilités et s'inspire de la formulation de Bruno de Finetti en terme de prix d'achat ou de vente d'un pari. Bruno de Finetti [18, 19] interprète la probabilité $P(A)$ d'un évènement $A \subseteq \Omega$ comme le prix qu'un agent est prêt à acheter ou vendre un pari qui rapporterait un gain d'une unité (par exemple 1 euro) si la réalisation de la variable incertaine X est dans A. Un tel pari peut donc être formulé comme la fonction caractéristique sur A.

Les paris de type fonctions caractéristiques, déjà suivant de Finetti, ne sont pas les seuls paris possibles sur la variable X. Un pari peut aussi être une fonction bornée sur Ω, $f : \Omega \to \mathbb{R}$ qui rapporterait $f(\omega)$ si la réalisation de la variable X est ω. On notera $\mathcal{L}(\Omega)$ l'ensemble des paris sur Ω. Les évènements A sont des cas particulier de pari $f = \mathbb{1}_A$. L'ensemble des paris à valeurs dans $\{0, 1\}$, i.e. l'ensemble des évènements, est noté $\mathcal{B}(\Omega)$, par analogie avec la tribu borélienne sur Ω.

Il est intéressant de remarquer que dans ce modèle des prévisions inférieures et supérieures de Peter Walley, Ω, l'ensemble référence, est toujours fini.

Dans ce modèle, la prévision inférieure, associée au pari f, est le prix maximum que l'agent (parieur) est prêt à mettre pour acheter le pari f. Cette prévision inférieure est notée $\underline{P}(f)$. La prévision supérieure, associée au pari f, est le prix minimum que l'agent (bookmaker) est prêt à vendre le pari f. Cette prévision supérieure est notée $\overline{P}(f)$. Ces deux prix sont liés. En effet, vendre un pari f à un prix s, c'est la même chose qu'acheter le pari $-f$ au prix $-s$. On lie donc le prix maximal d'achat au prix minimal de vente par la relation suivante :

$$\overline{P}(f) = -\underline{P}(-f). \tag{I.106}$$

De cette relation, on peut déduire que le modèle des prévisions inférieures est suffisant

pour représenter le modèle de Peter Walley car la prévision supérieure s'obtient à partir de la prévision inférieure par (I.106).

Évaluer la croyance sur les issues possibles d'une expérience par les prix qu'on est prêt à mettre pour acheter ou vendre des paris sur ces issues est finalement cohérent. Pour connaitre l'état de connaissance d'un agent rationnel et cohérent, observer sa façon de parier nous donne des indications sur ses croyances. Son attitude ou plutôt sa façon de parier doit être cohérente pour pouvoir définir un modèle mathématique de l'incertain qui ne donne pas de prévisions aberrantes. Si un parieur est prêt à payer au maximum 5 euros un pari f et si sur un pari qui rapporterait moins, i.e. un pari g tel que $\forall \omega \in \Omega$, $g(\omega) \leq f(\omega)$ il était prêt à payer plus, par exemple à payer au maximum 10 euros, il ne serait pas cohérent. Autrement dit,

$$f \geq g \Rightarrow \underline{P}(f) \geq \underline{P}(g). \tag{I.107}$$

Cette remarque est un corollaire des axiomes de cohérence de la théorie des prévisions inférieures de Peter Walley. Ces axiomes de cohérence, qui sont des extensions des axiomes de cohérence proposés par Bruno de Finetti dans le cas précis, sont les suivants :

Définition I.17 (Axiomes de cohérence) *Soient f et g des paris de $\mathcal{L}(\Omega)$. Soit λ un réel positif.*

1. $\underline{P}(f) \geq \inf_\Omega f(\omega)$ *acceptation des gains sûrs,*
2. $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$ *super additivité,*
3. $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ *homogénéité positive.*

Montrons que l'expression (I.107) est bien une conséquence des axiomes de cohérence.
Preuve : D'abord, comme $f \geq g$, il existe un pari positif h dans $\mathcal{L}(\Omega)$, tel que $f = g + h$. D'après l'axiome de super additivité, $\underline{P}(f) = \underline{P}(g+h) \geq \underline{P}(g) + \underline{P}(h)$. Or $\underline{P}(h) \geq 0$, car h (et donc sa borne inférieure) est positif. D'où $\underline{P}(f) \geq \underline{P}(g)$.

□

Dans la pratique, un agent ne peut définir de prévisions sur tous les paris possibles de Ω, i.e. pour tout $f \in \mathcal{L}(\Omega)$. L'extension naturelle, proposée par Peter Walley, c'est le principe qui permet d'étendre, de façon cohérente, les jugements qu'un agent effectue sur une famille de paris, à tous les paris sur Ω, i.e. à toutes les fonctions sur Ω réelles bornées. Ce procédé est particulièrement intéressant quand il s'agit d'étendre les jugements que l'agent a sur la famille des évènements $\mathcal{B}(\Omega)$ à des jugements sur la famille de toutes les fonctions bornées de Ω, i.e. à toute fonction de $f \in \mathcal{L}(\Omega)$. La définition de ce cas particulier d'extension naturelle est donnée :

Définition I.18 (Extension naturelle depuis $\mathcal{B}(\Omega)$) *Soit $\underline{P}$ une prévision inférieure (non nécessairement cohérente) sur $\mathcal{B}(\Omega)$. $\underline{E}$ l'extension naturelle de $\underline{P}$, pour tout $f \in \mathcal{L}(\Omega)$ est définie par :*

$$\underline{E}(f) = \sup\{\alpha : f - \alpha \geq \sum_{j=1}^{m} \lambda_j G(A_j),\ \textit{pour}\ m \geq 0, A_j \in \mathcal{B}(\Omega), \lambda_j \geq 0\}, \tag{I.108}$$

où $\forall A \in \mathcal{B}(\Omega)$, $G(A) = \mathbb{1}_A - \underline{P}(A)$.

Il est intéressant de rajouter un modèle équivalent à ce modèle de prévisions inférieures sur les paris. Ce modèle est basé sur les prévisions linéaires cohérentes. Les prévisions linéaires cohérentes, ce sont celles définis par Bruno de Finetti [18, 19], pour lequel le prix d'achat et le prix de vente sont les mêmes, autrement dit pour qui il existe une prévision P, telle que :

$$P(f) = \overline{P}(f) = \underline{P}(f). \tag{I.109}$$

Le modèle imprécis des prévisions inférieures cohérentes de Walley interprète l'imprécis comme un manque de connaissance sur la bonne prévision linéaire comme modèle de l'incertain. Notons $\mathbb{P}(\Omega)$ l'ensemble des prévisions linéaires sur Ω. On a ainsi qu'une prévision inférieure cohérente $\underline{P}$ est équivalente à une famille de prévisions linéaires par la relation :

$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\Omega) | \forall f \in \mathcal{L}(\Omega), P(f) \geq \underline{P}(f)\}. \tag{I.110}$$

De même, pour la prévision supérieure cohérente $\overline{P}$, associée à $\underline{P}$, elle est équivalente à une famille de prévisions linéaires par la relation :

$$\mathcal{M}(\overline{P}) = \{P \in \mathbb{P}(\Omega) | \forall f \in \mathcal{L}(\Omega), P(f) \leq \overline{P}(f)\} = \mathcal{M}(\underline{P}). \tag{I.111}$$

Ce modèle est intéressant car il permet une interprétation des prévisions inférieures cohérente comme un cadre qui permet de prendre en compte la méconnaissance sur le choix d'un modèle particulier de l'incertain. On se rapproche ici de l'imprécision telle que nous l'avons défini en section I.6.1, c'est-à-dire un modèle où l'imprécision serait représentée par un ensemble mesures de probabilités et donc un ensemble de noyaux sommatifs.

I.6.2.3 Les autres théories de l'incertain et de l'imprécis

La qualité et la complétude de l'étude de Peter Walley [105] sont très largement reconnues dans le monde des probabilités imprécises. Son modèle des prévisions inférieures cohérentes a cependant les défauts de ses qualités. C'est le modèle le plus général [105, 106, 107] des modèles de l'incertain et de l'imprécis. Et c'est donc souvent un modèle trop complexe, qui a besoin de trop de connaissances pour être évalué dans un problème. En l'occurrence il faut $2^{|\Omega|}$ évaluations de prix de paris pour que le modèle soit complet. Les modèles que nous présentons ci-après souffrent tous de la dualité suivante : plus le modèle est simple, moins il est expressif.

Voyons d'abord les probabilités inférieures et supérieures [105, 46]. C'est simplement la restriction du modèle précédent aux paris de type évènements, c'est-à-dire aux paris de $\mathcal{B}(\Omega)$. Les probabilités inférieures $\underline{P}$ et supérieures $\overline{P}$ sont également appelées mesures non-additives, mesures floues ou capacités [21], notées ν et vérifiant :

1. la propriété de monotonicité : $\forall A, B \in \mathcal{B}(\Omega)$, tels que $A \subseteq B$, on a $\nu(A) \leq \nu(B)$. Cette propriété est une version de la propriété (I.107) restreinte aux paris de types évènements,
2. les normalisations suivantes : $\nu(\emptyset) = 0$ et $\nu(\Omega) = 1$.

Le lien entre la probabilité inférieure et la probabilité supérieure d'un évènement A quelconque de $\mathcal{B}(\Omega)$ est maintenant donnée par :

$$\overline{P}(A) = 1 - \underline{P}(A^c). \tag{I.112}$$

Le terme mesure non-additive vient du fait que de telles mesures ne sont pas additives, i.e. que la mesure de l'union d'évènements disjoints n'est pas égale à la somme des mesures des évènements. Cette remarque étendue aux singletons permet de conclure qu'on ne peut pas, en général, exprimer une mesure non-additive à partir d'une distribution sur les singletons comme c'est le cas pour les mesures additives.

Une capacité de Choquet d'ordre 2 [14, 21, 45] est un cas particulier de probabilité inférieure, également appelée capacité convexe ou capacité 2-monotone.

Définition I.19. *Une probabilité inférieure $\underline{P}$ est une capacité convexe si elle est monotone d'ordre 2, i.e. si pour tous A et B de $\mathcal{B}(\Omega)$,*

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B). \tag{I.113}$$

Le conjugué $\overline{P}$ (I.112) d'une capacité convexe $\underline{P}$ est une capacité concave (propriété (I.113) avec inversion du sens de l'inégalité).

Une fonction de croyance [20, 92, 21] est un cas particulier de capacité convexe. Ce sont des capacités qui sont monotones d'ordre k , pour tout $k \in \mathbb{N}$,

Définition I.20. *Une probabilité inférieure $\underline{P}$ est une fonction de croyance si elle est k-monotone, pour tout $k \in \mathbb{N}$, i.e. si pour tous $(A_i)_{i=1,\ldots,k}$ de $\mathcal{B}(\Omega)$,*

$$\underline{P}(\bigcup_{i=1}^{k} A_i) + \sum_{I \subset \{1,\ldots,k\} \neq \emptyset} (-1)^{|I|} \underline{P}(\bigcap_{i \in I} A_i) \geq 0. \tag{I.114}$$

Le conjugué $\overline{P}$ (I.112) d'une fonction de croyance $\underline{P}$ est une fonction de plausibilité.

Une mesure de nécessité [30, 29, 116] est un cas particulier de fonction de croyance. Ceux sont des fonctions de croyance qui en plus d'être super additives (I.113) préservent le minimum.

Définition I.21. *Une probabilité inférieure $\underline{P}$ est une mesure de nécessité si elle préserve le minimum, i.e. si pour tous A et B de $\mathcal{B}(\Omega)$,*

$$\underline{P}(A \cap B) = min(\underline{P}(A), \underline{P}(B)). \tag{I.115}$$

Le conjugué $\overline{P}$ (I.112) d'une mesure de nécessité $\underline{P}$ est une mesure de possibilité.

Définition I.22. *Une probabilité supérieure $\overline{P}$ est une mesure de possibilité si pour tous A et B de $\mathcal{B}(\Omega)$,*

$$\overline{P}(A \cup B) = max(\overline{P}(A), \overline{P}(B)). \tag{I.116}$$

On notera Π, une mesure de possibilité et N, une mesure de nécessité.

Chapitre II

Extraction maxitive d'informations

Introduction

> *Ne faire qu'un petit bout de chemin n'est pas la même chose que se tromper de chemin. On ne se trompe pas en affirmant qu'Athènes est en Europe. Mais on ne peut pas dire que ce soit très précis non plus. Si un livre se contente d'écrire qu'Athènes est une ville située en Europe, il peut se révéler utile de consulter parallèlement un manuel de Géographie. Là tu apprendras toute la vérité, à savoir qu'Athènes est la capitale de la Grèce qui est un petit pays au sud-est de l'Europe. Avec un peu de chance, on parlera de l'Acropole aussi, voire de Socrate, Platon et Aristote !*

Nous vous proposons d'introduire notre approche sur l'extraction maxitive d'informations par cette citation, tirée de Le Monde de Sophie de *Jostein Gaarder*, publié en 1995 aux éditions Seuil. Ce roman est un cours sur la philosophie et son histoire faite à une adolescente de 14 ans.

La citation de Jostein Gaarder nous semble une bonne illustration de la façon dont l'Homme extrait de l'information sur des données, par le biais de connaissances. Dans cet exemple, les données c'est la réalité du monde, l'outil d'extraction d'informations ce sont les livres, les connaissances en général et enfin ce qu'on extrait, ce sont des informations imprécises sur la place d'Athènes dans le monde.

Relions dans un premier temps cette citation à l'extraction sommative d'informations : les données c'est la réalité, le noyau sommatif correspond au livre (appelons le ici "le livre sommatif"). La comparaison s'arrête ici, car l'information extraite est précise avec n'importe quel livre sommatif que l'on va utiliser. C'est ici une aberration de prétendre qu'avec n'importe quel livre, on peut avoir une information précise et correcte sur la place d'Athènes dans le monde. Cherchons la place d'Athènes dans Crimes et châtiments de *Fedor Dostoïevski*, il ne me semble pas pouvoir tirer de ce livre une information précise sur la localisation d'Athènes. Au mieux on pourra savoir que c'est en Grèce, et au pire on restera sur l'information qu'Athènes est sur la terre, qui est à l'échelle terrestre, l'information la plus imprécise qui soit.

Comme il est dit dans la citation, être imprécis et dans le vrai n'est pas la même chose que de se tromper précisément. Le cœur de ma thése, qu'est ce chapitre, est de développer un modèle imprécis d'extraction d'informations qui est plus sûrement dans le vrai que l'extraction sommative d'information précise. Dans notre approche, nous remplaçons le

livre sommatif par un autre outil de connaissances capable de mieux prendre en compte la méconnaissance, c'est la bibliothèque maxitive, mathématiquement représentée par le **noyau maxitif**.

Dans ces métaphores, nous prétendons qu'une bibliothèque prend mieux en compte la méconnaissance qu'un livre. Comment expliquer ce paradoxe ? Ceci est dû à la manière dont sont extraites les informations de ces objets livres ou bibliothèques. Avec un livre, on est trop optimiste et l'on s'acharne à extraire une information même si le livre ne convient pas. Avec la bibliothèque telle que nous proposons de l'utiliser, on ne va pas additionner, agréger ou fusionner les informations que l'on peut extraire de chacun des livres, on va simplement les regrouper comme un ensemble d'informations. On va rendre plus sûre notre information en la rendant moins précise.

Le modèle que nous proposons, sous le nom d'**extraction maxitive d'informations** est une méthodologie nouvelle de transformations de données en informations qui prend en compte la méconnaissance du modeleur sur le processus correct de transformation.

En section II.1, nous présentons le cadre général d'extraction imprécise d'informations. Cette présentation se veut être un parallèle de la présentation de l'extraction précise d'informations du chapitre précédent. La section II.2 est le point central de cette thèse où nous définissons la méthode d'extraction maxitive d'informations et le noyau maxitif.

Des liens entre le noyau maxitif et le noyau sommatif sont présentés en section II.3. Des questions relatives à ces liens sont posées ici, auxquelles nous tentons d'apporter des réponses justifiées comme le problème du choix d'un noyau maxitif ou de la séparabilité d'un noyau maxitif.

Nous justifions ensuite mathématiquement, en section II.4, la cohérence de l'extraction maxitive d'informations par rapport à l'extraction sommative d'informations. Nous terminons ce chapitre en proposant un indice du comportement des noyaux maxitifs en extraction maxitive d'informations en section II.5.

II.1 Extraction imprécise d'informations

Dans le chapitre précédent (cf. section I.1.3), nous avons présenté le cadre général de l'extraction précise d'informations, qui consiste à extraire une fonction d'informations y de $\mathbb{I}$ à partir d'une fonction de données x de $\mathbb{D}$, par le biais d'une transformation h. Rappelons que l'on parle d'extraction précise globale d'informations dans ce cas, car le résultat de la transformation h est une fonction d'information y définie pour tous les points θ de son ensemble référence Θ. On parle d'extraction précise locale d'information quand on utilise une transformation h^θ qui, pour une fonction de données x, renvoie une valeur $y(\theta)$ de la fonction d'information y en un unique point θ de Θ.

Dans cette section, nous introduisons le cadre théorique dans lequel nous souhaitons placer notre approche. Ce que nous présentons ici généralise l'extraction précise d'informations proposée au chapitre précédent en section I.1.3. Dans son principe, l'extraction imprécise d'informations consiste à extraire une famille convexe Y de fonctions d'informations y de $\mathbb{I}$ à partir d'une fonction de données x de $\mathbb{D}$, par le biais d'une transformation H. On notera $C(\mathbb{I})$ l'ensemble des familles convexes de fonctions d'informations de $\mathbb{I}$.

$$H : \begin{array}{ccc} \mathbb{D} & \rightarrow & C(\mathbb{I}), \\ x & \mapsto & Y = H(x), \end{array} \tag{II.1}$$

Le modèle d'extraction imprécise d'informations est dit global, car son résultat est un ensemble convexe de fonctions d'informations, qui sont toutes définies pour tous les θ de Θ. La figure II.1 présente un exemple qualitatif et arbitraire d'extraction imprécise globale d'informations. Le résultat de cette extraction imprécise globale d'informations est la famille convexe de fonctions d'informations $Y = H(x)$.

Parallèlement, le modèle local d'extraction imprécise d'information consiste à extraire un ensemble convexe $Y(\theta)$ de valeurs d'informations $y(\theta)$ en un point θ de Θ à partir d'une fonction de données x de $\mathbb{D}$, par le biais d'une transformation H^θ. On notera $C(\mathcal{Y})$ l'ensemble des ensembles convexes de $\mathcal{Y}$. On a :

$$H^\theta : \begin{array}{ccc} \mathbb{D} & \rightarrow & C(\mathcal{Y}), \\ x & \mapsto & Y(\theta) = H^\theta(x), \end{array} \tag{II.2}$$

La figure II.1 présente un exemple qualitatif et arbitraire d'extraction imprécise locale d'information en θ. Le résultat de cette extraction imprécise locale d'informations est l'ensemble convexe d'éléments d'informations $Y(\theta) = H^\theta(x)$.

Comme dans le cas sommatif, définir la transformation locale H^θ pour tout θ de Θ, revient à définir globalement l'extraction imprécise d'informations. En effet, pour une fonction de données quelconque $x \in \mathbb{D}$, $H(x)(\theta) = H^\theta(x)$, $\forall \theta \in \Theta$. De même, on constate de façon triviale que la convexité locale, i.e. la convexité de l'ensemble d'informations $Y(\theta)$, pour tout θ de Θ entraine la convexité globale de Y.

La convexité locale est décrite par $Y(\theta) \in C(\mathcal{Y}) \Leftrightarrow \forall y_1(\theta), y_2(\theta) \in Y(\theta)$, $\forall \lambda \in [0,1]$, $\lambda y_1(\theta) + (1-\lambda) y_2(\theta) \in Y(\theta)$. La convexité globale est décrite par $Y \in C(\mathbb{I}) \Leftrightarrow \forall y_1, y_2 \in Y$, $\forall \lambda \in [0,1]$, $\lambda y_1 + (1-\lambda) y_2 \in Y$.

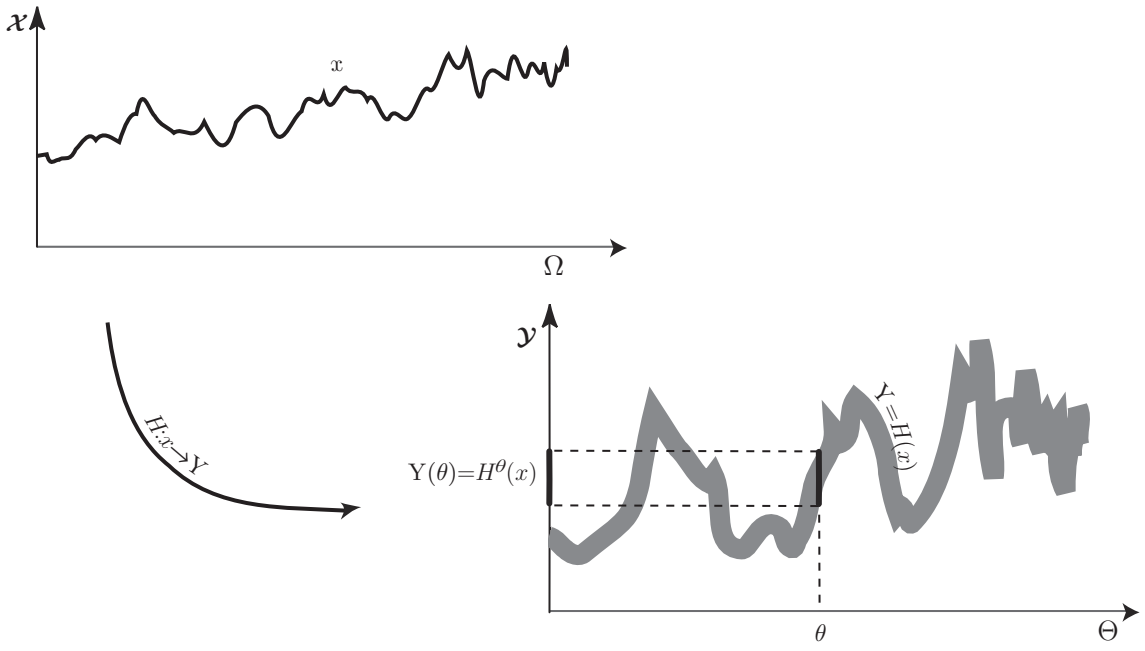


Fig. II.1 : Schéma général de l'extraction imprécise d'informations

II.2 Extraction maxitive d'informations

Le modèle d'extraction imprécise d'informations, tel que nous venons de le présenter, est très général. Son utilisation dans des applications nécessite le développement d'un ensemble de techniques, faciles à comprendre et à manipuler. C'est pourquoi nous avons mis au point une technique dont le principe se rapproche beaucoup de l'extraction sommative d'informations au sens où elle est aussi basée sur la définition d'un voisinage pondéré par noyau. Nous nous restreignons, pour ce modèle, à des données et des informations à valeurs réelles, i.e. $\mathcal{X} = \mathcal{Y} = \mathbb{R}$.

La clef du modèle d'extraction maxitive d'informations ne tient pas dans cette restriction aux réels, mais sur l'ajout de l'hypothèse que la transformation des données en un ensemble de valeurs d'informations en tout θ de Θ, i.e. que l'extraction imprécise locale d'informations, dépend de ce que nous convenons d'appeler un noyau maxitif.

II.2.1 Noyau maxitif

Définition II.1. *Un noyau maxitif π est une fonction définie sur Ω à valeur dans $[0,1]$, vérifiant la propriété de maxitivité :*

$$\sup_{\omega\in\Omega} \pi(\omega) = 1, \text{ dans le cas continu,} \tag{II.3}$$

$$\sup_{i\in\Omega} \pi_i = 1, \text{ dans le cas discret.} \tag{II.4}$$

Dans le cas discret, un noyau maxitif peut être considéré comme une suite $(\pi_i)_{i=1,\ldots,n}$ d'éléments de $[0,1]$ ou un vecteur de $[0,1]^n$.

Un noyau maxitif π peut être vu comme un voisinage pondéré (ou fenêtre pondérée) que l'on place sur le domaine Ω, et dont la magnitude correspond à des poids que l'on attribue à chaque élément ω de Ω.

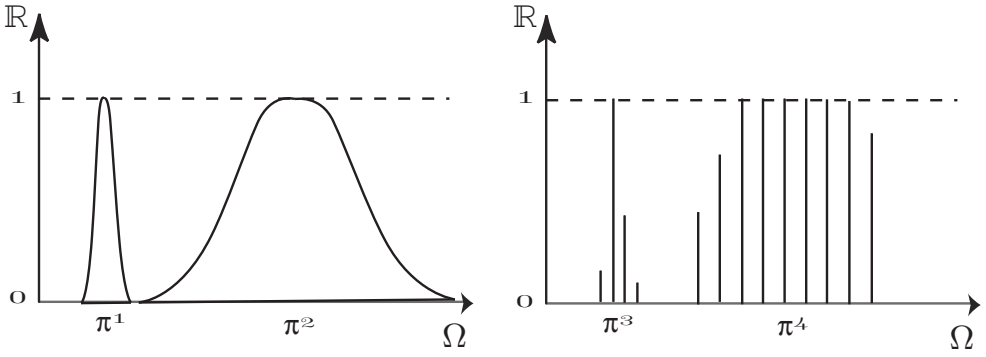


Fig. II.2 : Représentations qualitatives de noyaux maxitifs

De même qu'un noyau sommatif peut être interprété comme une distribution de probabilité, de même un noyau maxitif peut être interprété comme une distribution de possibilité. La définition d'un noyau maxitif coïncide avec la définition d'une distribution de possibilité [30]. Vu sous cet angle, un noyau maxitif π définit une mesure de possibilité,

notée Π_π, obtenue par :

$$\forall A \subseteq \Omega,\ \Pi_\pi(A) = \sup_{\omega \in A} \pi(\omega), \text{ dans le cas continu,} \tag{II.5}$$

$$\forall A \subseteq \Omega = \{1, ..., n\},\ \Pi_\pi(A) = \sup_{i \in A} \pi_i, \text{ dans le cas discret.} \tag{II.6}$$

La mesure de possibilité vérifie l'axiome de maxitivité suivant [30, 72] :

$$\forall (A_n)_{n>0} \subseteq \Omega,\ \Pi_\pi(\bigcup_{n>0} A_n) = \max_{n>0}(\Pi_\pi(A_n)). \tag{II.7}$$

Dans le cas discret, l'axiome suivant est suffisant :

$$\forall A, B \subseteq \Omega,\ \Pi_\pi(A \cup B) = \max(\Pi_\pi(A), \Pi_\pi(B)). \tag{II.8}$$

La théorie des possibilités, comme toutes les théories qui sont des cas particuliers de la théorie des prévisions inférieures et supérieures cohérentes de Peter Walley (cf. section I.6.2.2), fait intervenir deux bornes dans l'évaluation de l'incertain. La première est la mesure de possibilité (cf. expressions (II.7) et (II.8)), la seconde est la mesure de nécessité N_π, qui est le conjugué de Π_π, i.e.

$$\forall A \subseteq \Omega,\ N_\pi(A) = 1 - \Pi_\pi(A^c). \tag{II.9}$$

La mesure de nécessité peut être exprimée en fonction du noyau maxitif par :

$$\forall A \subseteq \Omega,\ N_\pi(A) = \inf_{\omega \notin A}(1 - \pi(\omega)), \text{ dans le cas continu,} \tag{II.10}$$

$$\forall A \subseteq \Omega = \{1, ..., n\},\ N_\pi(A) = \inf_{i \notin A}(1 - \pi_i), \text{ dans le cas discret.} \tag{II.11}$$

Une autre interprétation possible d'un noyau maxitif π, c'est le sous-ensemble flou. Depuis les travaux de Lotfi Zadeh [115, 116], on sait qu'une distribution de possibilité est identifiable à une fonction d'appartenance à un sous-ensemble flou normalisé. La fonction d'appartenance μ_F d'un sous-ensemble flou F est une fonction $\mu_F : \Omega \to [0, 1]$ (cf. définition I.11). On dit qu'un sous-ensemble flou est normalisé s'il existe au moins un $\omega \in \Omega$, tel que $\mu_F(\omega) = 1$.

Notons $\mathbb{M}_c(\Omega)$, l'ensemble des noyaux maxitifs continus sur Ω.

Notons $\mathbb{M}_d(\Omega)$, l'ensemble des noyaux maxitifs discrets sur Ω.

Notons $\mathbb{M}(\Omega)$, la réunion de ces ensembles, c'est-à-dire l'ensemble des noyaux maxitifs sur Ω.

II.2.2 Extraction maxitive d'informations

L'extraction maxitive d'informations est un cas particulier de l'extraction imprécise d'informations, définie en section II.1, dans lequel le noyau maxitif π^θ caractérise la façon dont sont agrégées les données $\{x(\omega)|\omega \in \Omega\}$ pour obtenir l'ensemble convexe des valeurs d'informations $Y(\theta)$.

Le noyau maxitif π^θ est défini sur Ω, l'ensemble référence des données. À chaque élément ω de Ω, correspondent une donnée $x(\omega)$ et un poids $\pi^\theta(\omega)$ attribué à $x(\omega)$ pour l'extraction d'information locale $Y(\theta)$. Dans ce cadre restreint de l'extraction maxitive d'informations, l'extraction proprement dite est réalisée par l'opérateur d'agrégation [65] qu'est l'intégrale de Choquet.

Définition II.2 (Intégrale de Choquet continue)
Soit ν une capacité sur Ω continu. Soit $x : \Omega \to \mathbb{R}$ une fonction bornée.

L'intégrale de Choquet de x par rapport à ν est définie par :

$$(C)\int_\Omega x d\nu = \int_{-\infty}^{0} (\nu(\{\omega \in \Omega | x(\omega) \geq \alpha\}) - 1)d\alpha + \int_0^{\infty} \nu(\{\omega \in \Omega | x(\omega) \geq \alpha\})d\alpha. \quad \text{(II.12)}$$

Définition II.3 (Intégrale de Choquet discrète)
Soit ν une capacité sur Ω discret, i.e. sur $\Omega = \{1, ..., n\}$. Soit $x = (x_i)_{i=1,...,n}$, un vecteur de $\mathbb{R}^n$.

L'intégrale de Choquet de x par rapport à ν est définie par :

$$(C)\int_\Omega x d\nu = \sum_{i=1}^{n} (x_{\sigma(i)} - x_{\sigma(i-1)})\nu(A_i), \quad \text{(II.13)}$$

où σ est la permutation sur Ω, telle que $x_{\sigma(1)} \leq ... \leq x_{\sigma(n)}$, avec $\sigma(0) = 0$ et $x_{\sigma(0)} = 0$ et où $A_i := \{\sigma(i), ..., \sigma(n)\}$.

Le passage de la forme continue à la forme discrète pour une intégrale de Choquet d'une fonction positive se fait par le calcul suivant :

Calcul II.4. *Supposons $x \geq 0$ discret, c'est-à-dire que x s'écrit comme une suite $(x_i)_{i=1,...n}$ d'éléments de $\mathbb{R}^+$. L'intégrale de Choquet devient alors, pour une capacité discrète :*

$$(C)\int_\Omega x d\nu = \int_0^{\infty} \nu(\{m \in \Omega | x_m \geq \alpha\})d\alpha.$$

On peut ordonner cette suite suivante un ordre croissant. On note σ la permutation sur x, telle que $x_{\sigma(1)} \leq ... \leq x_{\sigma(n)}$, avec $\sigma(0) = 0$ et $x_{\sigma(0)} = 0$. Ainsi, $\forall \alpha \in [0, \infty[$, $\exists i \in \{1, ..., n\}$, tel que $x_{\sigma(i-1)} < \alpha \leq x_{\sigma(i)}$. À un α donné, $\{m \in \Omega | x_m \geq \alpha\} = \{\sigma(i), ..., \sigma(n)\} = A_i$. D'où,

$$(C)\int_\Omega x d\nu = \sum_{i=1}^{n} \int_{x_{\sigma(i-1)}}^{x_{\sigma(i)}} \nu(A_i)d\alpha = \sum_{i=1}^{n} (x_{\sigma(i)} - x_{\sigma(i-1)})\nu(A_i).$$

□

Pour réaliser l'extraction maxitive d'informations à partir du noyau maxitif π^θ, l'opérateur intégrale de Choquet est utilisé deux fois. Cet opérateur est d'abord utilisé avec la mesure de nécessité N_{π^θ}, qui est un cas particulier de capacité convexe, pour obtenir la borne inférieure de l'intervalle d'extraction maxitive locale d'informations $Y(\theta)$. Il est ensuite utilisé avec la mesure de possibilité Π_{π^θ}, qui est un cas particulier de capacité concave, pour obtenir la borne supérieure de l'intervalle d'extraction maxitive locale d'informations $Y(\theta)$.

Par souci de simplification, on notera $\mathbb{C}_{\pi^\theta}(x)$, l'intégrale de Choquet, d'une fonction x bornée, par rapport à la mesure de possibilité Π_{π^θ} associée au noyau maxitif π^θ, i.e. $(C)\int_\Omega x d\Pi_{\pi^\theta} = \mathbb{C}_{\pi^\theta}(x)$. De même, on notera $\mathbb{C}^c_{\pi^\theta}(x)$, l'intégrale de Choquet, d'une fonction x bornée, par rapport à la mesure de nécessité N_{π^θ} associée au noyau maxitif π^θ, i.e. $(C)\int_\Omega x dN_{\pi^\theta} = \mathbb{C}^c_{\pi^\theta}(x)$.

Définition II.5 (Extraction maxitive d'information)
Sous les hypothèses :

1. *x est une fonction de données bornée,*
2. *$\pi^\theta \in \mathbb{M}(\Omega)$,*

l'extraction maxitive d'information en θ est obtenue par :

$$Y(\theta) = [\mathbb{C}^c_{\pi^\theta}(x), \mathbb{C}_{\pi^\theta}(x)]. \tag{II.14}$$

L'ensemble des informations locales extraite $Y(\theta)$ est convexe. En effet, comme nous nous sommes restreints à des noyaux maxitifs π^θ et des fonctions de données x à valeurs réelles, les opérations $\mathbb{C}^c_{\pi^\theta}(x)$ et $\mathbb{C}_{\pi^\theta}(x)$ sont à valeurs réelles. Un intervalle de $\mathbb{R}$ est un ensemble convexe, donc $Y(\theta)$ est convexe, pour tout θ de Θ. La convexité globale est directement déduite de la convexité locale, Y est donc convexe. Ces remarques montrent que l'extraction maxitive d'informations entre bien dans le cadre de l'extraction imprécise d'informations.

Dans le cas où $\Omega = \Theta$, c'est-à-dire où les données et les informations ont le même ensemble référence, il est possible de définir un noyau maxitif invariant par translation. Dans le cas continu, cela signifie qu'il existe un noyau maxitif de référence π, qui permet d'obtenir les noyaux maxitifs π^ω, pour tout ω de Ω.

$$\forall u \in \Omega,\ \pi^\omega(u) = \pi(\omega - u). \tag{II.15}$$

Dans le cas discret, sous l'hypothèse que l'on a un noyau maxitif π de référence invariant par translation, π^k, le translaté de π en $k \in \Omega$ est obtenu par :

$$\forall i \in \mathbb{Z},\ \pi^k_i = \pi_{k-i}. \tag{II.16}$$

Définir un noyau maxitif de référence (II.15) invariant par translation en tout point de Ω, nécessite que l'espace Ω soit stable par soustraction. Nous renvoyons à la section I.2.2 sur l'extraction sommative d'informations pour les remarques sur cette stabilité par soustraction.

On pourra également remarquer que chaque noyau maxitif continu, défini sur $\mathbb{R}$ peut être à l'origine d'une famille de noyaux maxitifs paramétrée par la largeur de bande $\Delta > 0$. Chaque élément de cette famille peut être obtenu par :

$$\pi_\Delta(\omega) = \pi(\frac{\omega}{\Delta}),\ \forall \omega \in \Omega. \tag{II.17}$$

Voici par exemple cinq éléments de la famille de noyaux triangulaires, obtenues par $\forall \omega \in \Omega$, $T_\Delta(\omega) = (1 - |\frac{\omega}{\Delta}|)\mathbb{1}_{|\omega| \leq \Delta}$.

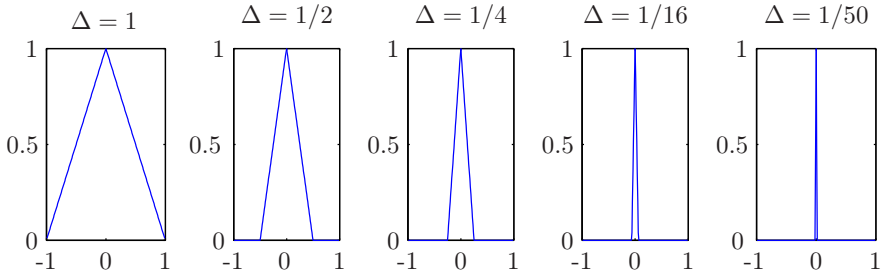


Fig. II.3 : Famille de noyaux maxitifs triangulaires

La largeur de bande est un facteur de dilatation du noyau κ_Δ. La largeur de bande est un facteur de dilatation du noyau π_Δ.

II.2.3 Extraction maxitive séquentielle d'informations

Pour mieux comprendre l'intérêt de l'extraction maxitive d'informations séquentielle, rappelons comment l'extraction sommative d'information séquentielle est formalisée sur une séquence de deux extractions consécutives. Dans un premier temps, on extrait de l'information y à partir des données de départ x avec les noyaux sommatifs κ^θ, définis pour tous θ de Θ. On a donc, d'après l'expression (I.14) que $\forall\theta \in \Theta$, $y(\theta) = \mathbb{E}_{\kappa^\theta}(x)$. Dans un deuxième temps, on extrait de l'information z à partir des données de départ y avec le noyau sommatif κ^ξ en ξ. On a donc, d'après l'expression (I.14) que $\forall\xi \in \Xi$, $z(\xi) = \mathbb{E}_{\kappa^\xi}(y)$.

L'extraction maxitive d'informations séquentielle est formalisée de la même manière. Dans un premier temps, on extrait de l'information Y à partir des données de départ x avec les noyaux maxitifs π^θ, définis pour tous θ de Θ. On a donc, d'après l'expression (II.14)

$$\forall\theta \in \Theta,\ Y(\theta) = [\mathbb{C}^c_{\pi^\theta}(x), \mathbb{C}_{\pi^\theta}(x)].$$

Dans un deuxième temps, on extrait de l'information Z à partir des données de départ y avec le noyau maxitif π^ξ en ξ. On a donc, d'après l'expression (II.14) que $\forall\xi \in \Xi$, $z(\xi) = \mathbb{E}_{\kappa^\xi}(y)$.

$$\forall\xi \in \Xi,\ Z(\xi) = [\mathbb{C}^c_{\pi^\xi}(\mathbb{C}^c_{\pi^\theta}(x)), \mathbb{C}_{\pi^\xi}(\mathbb{C}_{\pi^\theta}(x))].$$

II.3 Liens entre noyaux maxitifs et noyaux sommatifs

II.3.1 Normalisations et interprétation

Nous avons proposé une interprétation de voisinage pondéré pour le noyau sommatif et le noyau maxitif. La différence de normalisation de ces pondérations n'est pas très significative au premier abord. Pour définir un noyau sommatif ou maxitif, la donnée de départ est souvent un ensemble ou une suite de poids $d = (d(\omega))_{\omega\in\Omega}$ (en continu) et $d = (d_i)_{i=1,\dots,n}$ (en discret) de $\mathbb{R}^+$ non normalisée.

Cet ensemble de poids provient d'un savoir expert ou d'une analyse préalable du processus considéré.

Pour obtenir un noyau sommatif à partir de l'ensemble des poids d il suffit de diviser tous les poids par l'intégrale (dans le cas continu) ou la somme (dans le cas discret) des poids d sur Ω. Ainsi, d définit le noyau sommatif κ par la normalisation suivante :

$$\kappa(\omega) = \frac{d(\omega)}{\int_\Omega d(\omega)d\omega},\ \forall\omega \in \Omega, \text{ dans le cas continu,}$$
$$\kappa_i = \frac{d_i}{\sum_{l=1}^n d_l},\ \forall i \in \Omega, \text{ dans le cas discret.}$$

Pour obtenir un noyau maxitif à partir de l'ensemble des poids d il suffit de diviser tous les poids par la borne supérieure des poids d sur Ω. Ainsi, d définit le noyau maxitif π par la normalisation suivante :

$$\pi(\omega) = \frac{d(\omega)}{\sup_{\omega\in\Omega} d(\omega)},\ \forall\omega \in \Omega, \text{ dans le cas continu,}$$
$$\pi_i = \frac{d_i}{\sup_{l=1,\dots,n} d_l},\ \forall i \in \Omega, \text{ dans le cas discret.}$$

On pourrait s'interroger sur la signification de ces normalisations aux vues de leurs maigres différences. Ainsi, par construction, l'ordre des poids ne dépend pas du choix de la normalisation, $\forall \omega_1, \omega_2 \in \Omega$, si $d(\omega_1) \leq d(\omega_2)$, alors $\kappa(\omega_1) \leq \kappa(\omega_2)$ et $\pi(\omega_1) \leq \pi(\omega_2)$. Cette remarque est également vraie en discret avec l'écriture indicielle. Par ailleurs, le passage d'une normalisation à l'autre est mathématiquement très simple.

Pourtant la différence que le choix de normalisation implique dans l'interprétation du voisinage pondéré dans le monde des théories de l'incertain et de l'imprécis est essentielle pour le développement d'une extraction imprécise au lieu d'une extraction précise d'informations.

II.3.2 Noyau maxitif : spécificité et imprécision

Un noyau maxitif π, vu comme une distribution de possibilité, définit une mesure de possibilité Π_π par les relations (II.5) ou (II.6). Une mesure de possibilité est un cas particulier de probabilité supérieure cohérente $\overline{P}$ au sens de Walley définie pour les paris de $\mathcal{B}(\Omega)$ de type évènements. Or une probabilité supérieure cohérente $\overline{P}$ est équivalente à une famille $\mathcal{M}(\overline{P})$ de probabilités linéaires définies pour des évènements de $\mathcal{B}(\Omega)$, donc la mesure de possibilité Π_π est équivalente à la famille :

$$\mathcal{M}(\Pi_\pi) = \{P \in \mathbb{P}(\Omega) | \forall A \in \mathcal{B}(\Omega), P(A) \leq \Pi_\pi(A)\} \tag{II.18}$$

de probabilités linéaires. Les probabilités linéaires, qui sont des prévisions linéaires définies pour des évènements de $\mathcal{B}(\Omega)$, sont des mesures de probabilités. Donc Π_π, et donc le noyau maxitif π, est équivalent à la famille de mesures de probabilité $\mathcal{M}(\Pi_\pi)$. Une mesure de probabilité P est définie à partir d'une distribution de probabilité, qui peut être vue comme un noyau sommatif. On peut donc en conclure qu'un noyau maxitif π est équivalent à une famille de noyaux sommatifs, notée $\mathcal{M}(\pi)$ et définie par

$$\mathcal{M}(\pi) = \{\kappa \in \mathbb{S}(\Omega) | \forall A \in \mathcal{B}(\Omega), P_\kappa(A) \leq \Pi_\pi(A)\}. \tag{II.19}$$

Dans la suite de ce manuscrit, nous dirons qu'un noyau maxitif π domine un noyau sommatif κ ou qu'un noyau sommatif κ est dominé par un noyau maxitif π, si $\kappa \in \mathcal{M}(\pi)$, autrement dit, si $\forall A \in \mathcal{B}(\Omega)$, $P_\kappa(A) \leq \Pi_\pi(A)$.

L'imprécision induite par l'utilisation d'un noyau maxitif π est liée à la taille de la famille $\mathcal{M}(\pi)$ de noyaux sommatifs, c'est-à-dire au nombre de noyaux sommatifs dominés par π. Comment capturer cette imprécision dans un noyau maxitif ? La réponse n'est pas tout à fait directe. Il faut d'abord observer, comparer les imprécisions relatives de deux noyaux maxitifs. En particulier, on observe que lorsqu'un noyau maxitif π_1 est dominé, en terme de poids, par un noyau maxitif π_2, alors π_1 est plus précis que π_2 et inversement π_2 est plus imprécis que π_1. On note la domination pondérale $\succeq$, définie par $\pi_2 \succeq \pi_1 \Leftrightarrow \forall \omega \in \Omega$, $\pi_1(\omega) \leq \pi_2(\omega)$. On montre facilement que $\pi_2 \succeq \pi_1 \Rightarrow \mathcal{M}(\pi_1) \subseteq \mathcal{M}(\pi_2)$.
Preuve : Supposons que $\pi_2 \succeq \pi_1$, alors :

$$\Pi_{\pi_1} \leq \Pi_{\pi_2}, \tag{II.20}$$
$$N_{\pi_2} \leq N_{\pi_1}. \tag{II.21}$$

Autrement dit,

$$N_{\pi_2} \leq N_{\pi_1} \leq \Pi_{\pi_1} \leq \Pi_{\pi_2}. \tag{II.22}$$

Ces inégalités sont triviales à partir des définitions des mesures de possibilité et nécessité.

Alors, tout noyau sommatif dans $\mathcal{M}(\pi_1)$ est dans $\mathcal{M}(\pi_2)$, donc $\mathcal{M}(\pi_1) \subseteq \mathcal{M}(\pi_2)$.

□

Précisons ici que la domination pondérale d'un noyau maxitif π_2 sur un autre π_1 entraîne la domination en mesure de π_2 sur π_1, i.e. $\forall A \in \mathcal{B}(\Omega)$, $\Pi_{\pi_1}(A) \leq \Pi_{\pi_2}(A)$.

La notion de spécificité des distributions de possibilité ou des sous-ensembles flous est aussi exprimée par comparaison des spécificités relatives de deux distributions de possibilité, qui s'appuie également sur une comparaison des poids des distributions de possibilité utilisées. On dit qu'un noyau maxitif π_1 est plus spécifique qu'un noyau maxitif π_2, si $\forall \omega \in \Omega$, $\pi_1(\omega) \leq \pi_2(\omega)$. On observe directement que la notion de spécificité des distributions de possibilité est inversement liée à la notion d'imprécision du noyau maxitif.

L'ordre de spécificité entre noyaux maxitifs va également classer les noyaux maxitifs par ordre de précision. Plus un noyau maxitif est spécifique, plus il est précis.

La spécificité telle que présentée ici pour un noyau maxitif, ou une distribution de possibilité, a son homologue dans la théorie des sous-ensembles flous. Dans cette interprétation du noyau maxitif, la spécificité est 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 dans le cas continu et du cardinal de l'ensemble dans le cas discret.

La notion de spécificité est donc commune aux noyaux sommatifs et maxitifs. Il s'agit de la concentration des poids du noyau considéré sur un ensemble de longueur minimale. Dans le chapitre précédent, nous avons proposé la spécificité d'un noyau sommatif comme caractéristique de son comportement en extraction sommative d'informations. Maintenant, nous voyons que la spécificité d'un noyau maxitif caractérise sa précision.

II.3.3 Séparabilité des noyaux maxitifs

Rappelons d'abord qu'un noyau sommatif κ_{12}, défini sur un produit cartésien d'ensembles, noté $\Omega_1 \times \Omega_2$, est dit séparable, quand il peut s'exprimer comme le produit de noyaux sommatifs κ_1 sur Ω_1 et κ_2 sur Ω_2 (cf. définition I.5). C'est-à-dire :

$$\forall(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2,\ \kappa_{12}(\omega_1, \omega_2) = \kappa_1(\omega_1)\kappa_2(\omega_2). \tag{II.23}$$

Pour la même raison, c'est-à-dire pour une simplicité des calculs, on propose les définitions de séparabilité suivante pour les noyaux masitifs :

Définition II.6 (prod-séparabilité)
Le noyau maxitif $\pi_{1\times2}$, défini $\Omega_1 \times \Omega_2$, est dit prod-séparable, quand il peut s'exprimer comme le produit des noyaux maxitifs π_1 sur Ω_1 et π_2 sur Ω_2 :

$$\forall(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2,\ \pi_{1\times2}(\omega_1, \omega_2) = \pi_1(\omega_1)\pi_2(\omega_2). \tag{II.24}$$

Définition II.7 (min-séparabilité)
Le noyau maxitif $\pi_{1\wedge2}$, défini $\Omega_1 \times \Omega_2$, est dit min-séparable, quand il peut s'exprimer, à partir des noyaux maxitifs π_1 sur Ω_1 et π_2 sur Ω_2, par :

$$\forall(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2,\ \pi_{1\wedge2}(\omega_1, \omega_2) = min(\pi_1(\omega_1), \pi_2(\omega_2)). \tag{II.25}$$

Théorème II.8. *Soient π_1 et π_2 deux noyaux maxitifs respectivement définis sur Ω_1 et Ω_2. Pour tous noyaux sommatifs $\kappa_1 \in \mathcal{M}(\pi_1)$ et $\kappa_2 \in \mathcal{M}(\pi_2)$, le noyau sommatif séparable κ_{12} construit par (I.25) est dominé par $\pi_{1\times 2}$, construit par (II.24), lui même dominé par $\pi_{1\wedge 2}$, construit par (II.25). Ces dominations sont valables pour des évènements de type pavé : $A = A_1 \times A_2$ tels que $A_1 \subseteq \Omega_1$ et $A_2 \subseteq \Omega_2$.*

Preuve : On constate d'abord que $\pi_{1\wedge 2}$ domine $\pi_{1\times 2}$, car $\forall(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, $\pi_1(\omega_1) \leq 1$ et $\pi_2(\omega_2) \leq 1$, donc $\pi_1(\omega_1)\pi_2(\omega_2) \leq \pi_1(\omega_1)$, mais également $\pi_1(\omega_1)\pi_2(\omega_2) \leq \pi_2(\omega_2)$, donc $\pi_1(\omega_1)\pi_2(\omega_2) \leq min(\pi_1(\omega_1), \pi_2(\omega_2))$. La domination en mesure est alors triviale, i.e. $\forall A = A_1 \times A_2$ tels que $A_1 \subseteq \Omega_1$ et $A_2 \subseteq \Omega_2$, $\Pi_{\pi_{1\wedge 2}}(A) \geq \Pi_{\pi_{1\times 2}}(A)$.

Il suffit maintenant de montrer que $\pi_{1\times 2}$ domine κ_{12}, autrement dit $\forall A = A_1 \times A_2$ tels que $A_1 \subseteq \Omega_1$ et $A_2 \subseteq \Omega_2$, $\Pi_{\pi_{1\times 2}}(A) \geq P_{\kappa_{12}}(A)$.

$$\begin{aligned}
\Pi_{\pi_{1\times 2}}(A) &= \sup_{\omega\in A} \pi_{1\times 2}(\omega), \\
&= \sup_{\omega_1\in A_1, \omega_2\in A_2} \pi_1(\omega_1)\pi_2(\omega_2), \\
&= \{\sup_{\omega_1\in A_1} \pi_1(\omega_1)\}\{\sup_{\omega_2\in A_2} \pi_2(\omega_2)\}, \\
&= \Pi_{\pi_1}(A_1)\Pi_{\pi_2}(A_2). \\
P_{\kappa_{12}}(A) &= \int_A \kappa_{12}(\omega)d\omega, \\
&= \int_{A_1}\int_{A_2} \kappa_1(\omega_1)\kappa_2(\omega_2)d\omega_2 d\omega_1, \\
&= \int_{A_1} \kappa_1(\omega_1)d\omega_1 \int_{A_2} \kappa_2(\omega_2)d\omega_2 \\
&= P_{\kappa_1}(A_1)P_{\kappa_2}(A_2).
\end{aligned}$$

Puisque $\Pi_{\pi_1}(A_1) \geq P_{\kappa_1}(A_1)$ et $\Pi_{\pi_2}(A_2) \geq P_{\kappa_2}(A_2)$, on a

$$\Pi_{\pi_{1\times 2}}(A) \geq P_{\kappa_{12}}(A). \tag{II.26}$$

□

II.3.4 Le choix du noyau maxitif

Malgré une prolifération de la théorie des possibilités ces dernières années en traitement des données statistiques ou traitement du signal, les utilisateurs ont toujours tendance à proposer des méthodes de choix de noyau maxitif basées sur une simple renormalisation possibiliste d'un noyau sommatif (comme présentée en section II.3.1). La plupart du temps, le noyau sommatif original est choisi par habitude. Par exemple les statisticiens privilégieront le noyau d'Epanechnikov, les spécialistes du traitement d'images, les noyaux gaussien, uniforme ou même pour certains les B-splines. Le choix d'un noyau maxitif, avec l'idée sous-jacente qu'il représente une famille de modèles sommatifs, n'est pas encore rentré dans les mœurs. C'est sur cette idée que nous proposons des méthodes de choix de noyau maxitifs basées sur les noyaux sommatifs.

II.3.4.1 Le noyau maxitif triangulaire symétrique

L'utilisateur peut émettre l'hypothèse que le modèle précis ou sommatif sous-jacent à l'extraction locale d'information en θ_0 est forcément un noyau sommatif symétrique centré en un point ω_0 de l'ensemble référence des données $\Omega \subseteq \mathbb{R}$, continu et de support $[\omega_0 - \Delta, \omega_0 + \Delta]$. Cette hypothèse est très courante en extraction d'information, c'est le cas par exemple pour la méthode à noyaux de Parzen Rozenblatt ou pour les passages C/D en traitement du signal, où l'agrégation de données se fait avec une fenêtre pondérée symétrique. Dans ce cas, c'est l'utilisation du noyau maxitif triangulaire qui est le plus judicieux. En effet, Dubois et al. [31] ont montré que tout noyau sommatif symétrique centré en ω_0 et de support $[\omega_0-\Delta, \omega_0+\Delta]$ est dominé par le noyau maxitif triangulaire $T_\Delta^{\omega_0}$, centré en ω_0 et de support $[\omega_0 - \Delta, \omega_0 + \Delta]$. En termes de familles de noyaux sommatifs, on a le lien suivant : si on note $\mathbb{S}_c^{s\omega_0\Delta}(\Omega)$, l'ensemble des noyaux symétriques centrés en ω_0 et de support $[\omega_0 - \Delta, \omega_0 + \Delta]$, on a $\mathbb{S}_c^{s\omega_0\Delta}(\Omega) \subset \mathcal{M}(T_\Delta^{\omega_0})$.

On remarque également que le choix du maxitif triangulaire est judicieux dans le cas d'extraction maxitive d'informations séquentielle. Supposons que l'on réalise successivement des extractions maxitives avec la famille de noyaux triangulaires $(T_{\Delta'}^u)_{u\in\Omega}$, de supports $[u - \Delta', u + \Delta']$, puis avec le noyau maxitif T_Δ^ω de support $[\omega - \Delta, \omega + \Delta]$.

Dubois et al. [31] ont montré que pour tous les noyaux sommatifs $(\kappa^u)_{u\in\Omega}$ et κ^ω dominés respectivement par $(T_{\Delta'}^u)_{u\in\Omega}$ et T_Δ^ω, la combinaison convolutive de deux noyaux sommatifs κ^u et κ^ω, définie pour tout $v \in \Omega$, par $\eta^\omega(v) = \int_\Omega \kappa^u(v)\kappa^\omega(u)du$ (cf. section I.24) était dominé par la combinaison sup-min de deux noyaux triangulaires $(T_{\Delta'}^u)_{u\in\Omega}$ et T_Δ^ω.

La combinaison sup-min de deux noyaux triangulaires $(T_{\Delta'}^u)_{u\in\Omega}$ et T_Δ^ω est équivalente à la somme floue [30, 115, 117] des sous-ensembles flous définis par les noyaux maxitifs $(T_{\Delta'}^u)_{u\in\Omega}$ et T_Δ^ω.

La combinaison sup-min des noyaux triangulaires $(T_{\Delta'}^u)_{u\in\Omega}$ avec le noyau triangulaire T_Δ^ω est donnée, pour tout $v \in \Omega$, par :

$$T_{\Delta+\Delta'}^\omega(v) = \sup_{u\in\Omega} \min(T_{\Delta'}^u(v), T_\Delta^\omega(u)), \tag{II.27}$$

Le noyau $T_{\Delta+\Delta'}^\omega$ est un noyau maxitif de support $[\omega - (\Delta + \Delta'), \omega + (\Delta + \Delta')]$.

Tous les noyaux sommatifs de convolution η^ω sont dominés par le noyau triangulaire sup-min $T_{\Delta+\Delta'}^\omega$. Ce noyau maxitif triangulaire n'est certainement pas le plus spécifique des noyaux maxitifs qui dominent toutes les combinaisons convolutives des noyaux sommatifs κ^u et κ^ω, car certains noyaux sommatifs symétriques dominés par $T_{\Delta+\Delta'}^\omega$ ne sont pas obligatoirement décomposables en une convolution de deux noyaux symétriques κ^u et κ^ω dominés respectivement par $T_{\Delta'}^u$ et T_Δ^ω.

Cette approche est cependant très simple pour un choix de noyau maxitif synthétisant une séquence d'extractions maxitives d'informations. Ainsi, dans le cas où l'on réalise une séquence de m extractions maxitives d'informations séquentielle avec des noyaux triangulaires de largeurs de support successifs $\Delta_1, \Delta_2, ..., \Delta_m$, l'utilisation du noyau triangulaire $T_{\sum_{j=1}^m \Delta_j}^\omega$ peut être proposée pour réaliser une seule extraction maxitive d'informations au lieu de m extractions successives.

II.3.4.2 Inégalité de Chebychev

Dans le même ordre d'idée, on peut utiliser des notions de la théorie des probabilités, comme des inégalités, pour dégager une famille de noyaux sommatifs. Dubois et al [31, 34]

proposent notamment l'utilisation de l'inégalité de Chebychev, qui permet de définir un noyau maxitif dominant tous les noyaux sommatifs, dont la moyenne m et la variance σ sont données. Ce noyau maxitif, noté $\pi_{(m,\sigma)}$ est symétrique. Il est défini sur un ensemble référence $\Omega = \mathbb{R}$ par :

$$\pi_{(m,\sigma)}(m - \omega\sigma) = \pi_{(m,\sigma)}(m + \omega\sigma) = \min(1, \frac{1}{\omega^2}),\ \forall\omega > 0.$$

On note $\mathbb{S}_c^{(m,\sigma)}(\mathbb{R})$, l'ensemble des noyaux sommatifs de moyenne m et de variance σ. Ainsi, si l'utilisateur émet l'hypothèse que le modèle précis ou sommatif sous-jacent à l'extraction locale d'information en θ_0 est forcément un noyau sommatif de moyenne m et de variance σ, il pourra utiliser le noyau maxitif $\pi^{\omega_0}_{(m,\sigma)}$. On a cette fois que $\mathbb{S}_c^{(m,\sigma)}(\mathbb{R}) \subset \mathcal{M}(\pi^{\omega_0}_{(m,\sigma)})$. Cependant, cette famille est très peu spécifique : elle contient beaucoup de noyaux sommatifs qui ne sont pas nécessairement de moyenne m et de variance σ.

II.3.4.3 Transformations d'un noyau sommatif en noyau maxitif

Voyons une autre piste : si l'utilisateur a une idée du noyau sommatif κ à utiliser, nous pouvons proposer d'utiliser un noyau maxitif issue d'une transformation de ce noyau κ. Une telle transformation peut s'avérer utile pour quelqu'un qui ne maîtrise pas les fondements de la théorie des possibilités mais qui a l'habitude d'utiliser un noyau sommatif. Deux transformations sont présentées ici. La première est dite objective et la seconde est dite subjective.

II.3.4.3.1 Transformation objective

Dans l'interprétation objectiviste des probabilités, un intervalle de confiance de niveau de confiance α du noyau sommatif κ, est un sous ensemble de Ω, tel que le degré de probabilité, pour toute réalisation du phénomène incertain sous-jacent, de tomber dans ce sous ensemble soit égal à α. Si on note I_α un intervalle de confiance de niveau α, on a $P_\kappa(I_\alpha) = \alpha$. Ces intervalles de confiance sont la clef de la transformation objective de Dubois et al. [32, 31, 34]. Les coupes de niveau $(1 - \alpha)$ de la distribution de possibilité résultant de cette transformation sont les intervalles de confiance de niveau α les plus spécifiques de la distribution de probabilité κ à transformer. Les intervalles de confiance de niveau α les plus spécifiques de la distribution de probabilité κ sont les plus petits intervalles, au sens de leur longueur de Lebesgue λ, vérifiant $P_\kappa(I_\alpha) = \alpha$. On retrouve ici la notion de spécificité évoquée dans les sections I.5 et II.3.2.

Il a été montré que le résultat de cette transformation, que nous notons $\pi_{\leftarrow\kappa}$, est la distribution de possibilité la plus spécifique des distributions de possibilités qui vérifient les conditions intuitives suivantes :

- préservation de l'ordre des poids : $\kappa(\omega_1) \geq \kappa(\omega_2) \Rightarrow \pi_{\leftarrow\kappa}(\omega_1) \geq \pi_{\leftarrow\kappa}(\omega_2)$,
- principe de domination : $\kappa \in \mathcal{M}(\pi_{\leftarrow\kappa})$.

Nous donnons ici une interprétation différente à cette transformation en détaillant les étapes d'une construction différente. Nous interprétons le résultat de cette transformation $\pi_{\leftarrow\kappa}$, comme un noyau maxitif résumant les informations de spécificité du noyau sommatif κ.

Comme nous l'avons déjà évoqué en section I.5, la spécificité d'un noyau sommatif est sa capacité à être concentré sur un ensemble de longueur minimale. Or les intervalles de

confiance de niveau α permettent de comparer la spécificité, au niveau α, de deux noyaux sommatifs.

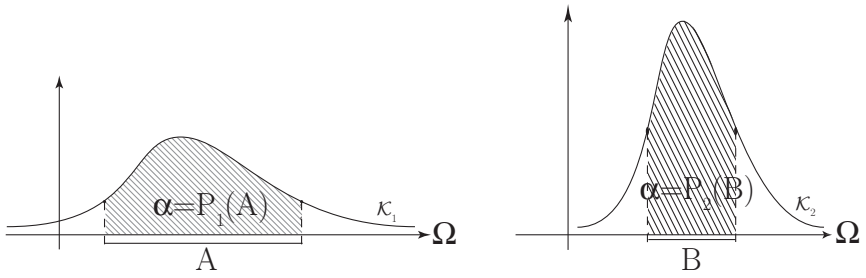


Fig. II.4 : Comparaison de spécificité de niveau α de κ_1 et κ_2

La figure II.4 est composée de deux noyaux sommatifs κ_1 et κ_2 dessinés qualitativement. Il est intuitif de comparer, à un niveau de confiance α donné, les intervalles de confiance de niveau α les plus spécifiques pour comparer la spécificité relative au niveau α de κ_1 et κ_2. Pour κ_1, l'intervalle de confiance le plus spécifique est A, pour κ_2, l'intervalle de confiance le plus spécifique est B. On remarque que B est plus spécifique que A, c'est-à-dire que $\lambda(B) < \lambda(A)$, λ étant la mesure de Lebesgue. Donc κ_2 est plus spécifique que κ_1 au niveau α. Comment maintenant résumer ou agréger les informations sur les spécificités de niveaux α, pour tous $\alpha \in [0, 1]$?

Pour cela, nous considérons les intervalles de confiance non plus pour des niveaux α, mais pour chacun des éléments du domaine de définition du noyau sommatif, i.e. pour tout ω de Ω. Ces intervalles de confiance sont définis, pour tout ω de Ω, par :

$$I_\omega = \{x \in \Omega / \kappa(x) \geq \kappa(\omega)\}. \tag{II.28}$$

I_ω est un intervalle de confiance de niveau $P_\kappa(I_\omega)$. On peut remarquer que les intervalles $(I_\omega)_{\omega\in\Omega}$ forment une famille d'intervalles emboîtés. En effet, pour tous ω_1 et ω_2 de Ω, soit $\kappa(\omega_2) \geq \kappa(\omega_1)$, on a alors $I_{\omega_2} \subseteq I_{\omega_1}$, soit $\kappa(\omega_1) \geq \kappa(\omega_2)$, et dans ce cas, on a $I_{\omega_1} \subseteq I_{\omega_2}$.

L'ensemble des intervalles de confiance $(I_\omega)_{\omega\in\Omega}$ permet d'ordonner les éléments ω de Ω en fonction de leur importance dans la quantification de la spécificité de κ. En effet, plus I_ω est spécifique (au sens de sa longueur), plus ω a du poids dans la quantification de la spécificité de κ. La figure II.5 illustre qualitativement cette conjecture. Elle montre deux intervalles de confiances I_{ω_1} et I_{ω_2} du noyau sommatif κ, tels que I_{ω_2} est plus spécifique que I_{ω_1}.

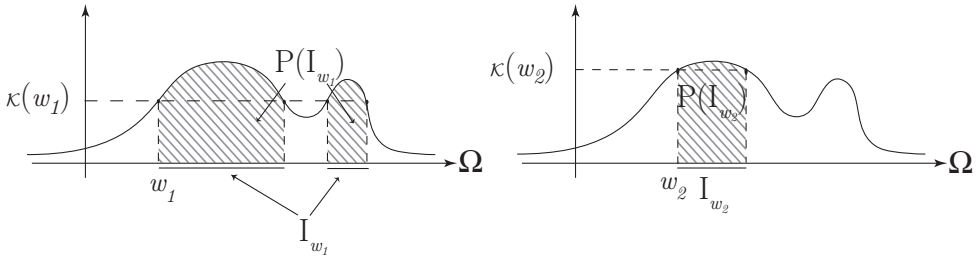


Fig. II.5 : Degré d'implication de ω_1 et ω_2 dans la spécificité de κ

Si l'on cherche à classer les éléments ω de Ω suivant leur apport à la spécificité totale de κ, on peut dire que ω_2 a un degré d'implication plus important dans la spécificité de

κ que ω_1. Autrement dit, $\lambda(I_{\omega_2}) < \lambda(I_{\omega_1})$. L'emboîtement des intervalles implique que $I_{\omega_2} \subseteq I_{\omega_1}$ et donc $P_\kappa(I_{\omega_2}) < P_\kappa(I_{\omega_1})$. Un ordre entre les éléments ω de Ω, qui correspond à leur poids dans la spécificité de κ, peut être donné par :

$$\omega_1 \preceq \omega_2 \Longleftrightarrow P_\kappa(I_{\omega_1}) \geq P_\kappa(I_{\omega_2}).$$

Cette ordre est équivalent à :

$$\omega_1 \preceq_{sp} \omega_2 \Longleftrightarrow 1 - P_\kappa(I_{\omega_1}) \leq 1 - P_\kappa(I_{\omega_2}). \tag{II.29}$$

Les valeurs $(1 - P_\kappa(I_\omega))$ atteignent leur maximum au mode de κ ; ce maximum est égal à 1. Ces valeurs ont donc une normalisation maxitive, et nous proposons comme noyau maxitif des poids de spécificité de κ, le noyau suivant :

$$\forall \omega \in \Omega,\ \pi_{\leftarrow\kappa}(\omega) = 1 - P_\kappa(I_\omega). \tag{II.30}$$

Ce noyau maxitif résume donc les informations de spécificité du noyau sommatif κ, et c'est le plus spécifique des noyaux maxitifs qui respectent les conditions de préservation de l'ordre des poids et de domination de $\pi_{\leftarrow\kappa}$ sur κ.

II.3.4.3.2 Transformation subjective

Voyons maintenant une transformation moins spécifique que $\pi_{\leftarrow\kappa}$ qui respecte ces mêmes conditions. C'est la transformation inverse à la transformation pignistique de Philippe Smets [98, 33, 36, 32], qui transforme une fonction de croyance $\underline{P}$ en la mesure de probabilité P, qui est le centre de gravité de $\mathcal{M}(\underline{P})$. Le fait de qualifier cette transformation de subjective, vient de la nature subjective de la théorie des fonctions de croyance transférables de Philippe Smets [98]. La transformation subjective est définie, dans le cas où Ω est continu, par :

Définition II.9. *Soit κ un noyau sommatif défini sur Ω. Le noyau maxitif $\pi_{[\kappa]}$, obtenue par la transformation subjective de κ, est défini par :*

$$\forall \omega \in \Omega, \pi_{[\kappa]}(\omega) = \int_\Omega \min(\kappa(x), \kappa(\omega))dx. \tag{II.31}$$

Dans le cas discret, l'expression (II.31) devient

$$\forall i \in \{1, ..., n\}, \pi_{[\kappa]i} = \sum_{j=i}^{n} \min(\kappa_i, \kappa_j). \tag{II.32}$$

Il est évident, quand on observe l'expression (II.31), que $\pi_{[\kappa]}(\omega)$ est égale à l'aire de la partie hachurée $(a) + (b)$ de la figure II.6, avec $(a) = \pi_{\leftarrow\kappa}(\omega)$ et $(b) = \lambda(I_\omega)\kappa(\omega)$.

Théorème II.10. *Soit κ un noyau sommatif défini sur Ω. Pour tout $\omega \in \Omega$, on a :*

$$\pi_{[\kappa]}(\omega) = \pi_{\leftarrow\kappa}(\omega) + \lambda(I_\omega)\kappa(\omega). \tag{II.33}$$

La différence en ω entre $\pi_{[\kappa]}(\omega)$ et $\pi_{\leftarrow\kappa}(\omega)$, qui est $(b) = \lambda(I_\omega)\kappa(\omega)$, est positive, donc la transformation subjective est moins spécifique que la transformation objective, i.e. $\pi_{[\kappa]} \preceq_{SP} \pi_{\leftarrow\kappa}$.

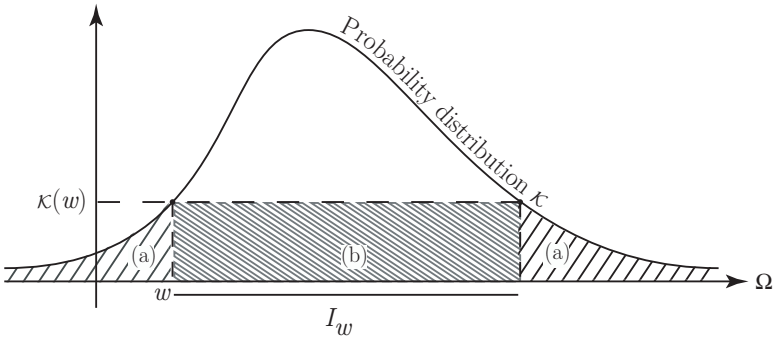


Fig. II.6 : illustration des transformations $\pi_{[\kappa]}(\omega)$ et $\pi_{\leftarrow\kappa}(\omega)$

II.3.4.3.3 Aide aux praticiens

Le tableau II.1 est une liste de noyaux sommatifs symétriques centrés, définis sur $\mathbb{R}$, souvent utilisés en traitement du signal ou en statistique, associés à leurs transformations objective et subjective.

	$u(\omega)$	$v(\omega)$	$t(\omega)$
	$D_u = [-1, 0]$	$D_v = [-1, 0]$	$D_t = [-1, 0]$
uniforme	$1/2$	$1 + \omega$	$1 + 2\omega$
triangulaire	$(1 + \omega)$	$1 + 2\omega + \omega^2$	$1 + 4\omega + 3\omega^2$
Epanechnikov	$\frac{3}{4}(1 - \omega^2)$	$1 + \frac{3}{2}\omega - \frac{\omega^3}{2}$	$1 + 3\omega - 2\omega^3$
triweight	$\frac{35}{32}(1 - \omega^2)^3$	$1 - \frac{35}{16}(\omega - \omega^3 + \frac{3\omega^5}{5} - \frac{\omega^7}{7})$	$1 - \frac{35}{16}(2\omega^3 - \frac{12\omega^5}{5} + \frac{6\omega^7}{7})$
cosinus	$\frac{\pi}{4}\cos(\frac{\pi}{2}\omega)$	$1 + \sin(\frac{\pi}{2}\omega)$	$1 + \sin(\frac{\pi}{2}\omega) + \frac{\pi}{2}\omega\cos(\frac{\pi}{2}\omega)$
	$D_u =]-\infty, 0]$	$D_v =]-\infty, 0]$	$D_t =]-\infty, 0]$
exponentiel	$\frac{1}{2}e^{\omega}$	e^{ω}	$e^{\omega}(1 + \omega)$

Tab. II.1 : Transformations de noyaux sommatifs usuels

La première colonne contient la fonction u, qui est la partie négative du noyau sommatif κ, que l'on obtient par :

$$\kappa(\omega) = u(-|\omega|). \tag{II.34}$$

La seconde colonne contient la fonction v, qui est la partie négative du noyau maxitif $\pi_{\leftarrow\kappa}$, que l'on obtient par :

$$\pi_{\leftarrow\kappa}(\omega) = v(-|\omega|). \tag{II.35}$$

La troisième colonne contient la fonction t, qui est la partie négative du noyau maxitif $\pi_{[\kappa_\Delta]}$, que l'on obtient par :

$$\pi_{[\kappa]}(\omega) = t(-|\omega|). \tag{II.36}$$

Ce tableau peut être particulièrement utile à un praticien du traitement des données.

II.4 Cohérence de l'extraction maxitive d'informations

La méthode d'extraction sommative d'informations est très répandue en traitement du signal et des images ainsi qu'en statistiques, comme nous l'avons montré au chapitre

I. Cette large utilisation devient un avantage pour la généralisation que nous proposons au travers de l'extraction maxitive d'informations. Les liens que nous avons dégagé entre les noyaux maxitifs et les noyaux sommatifs à la section II.3 justifient le choix de l'outil noyau maxitif dans une finalité d'amélioration de la robustesse de l'extraction sommative d'informations.

Ce qui manque encore comme justification, c'est l'étude et la comparaison des résultats obtenus par ces deux méthodes d'extraction d'informations. Nous regardons dans cette section la cohérence entre le résultat d'une extraction maxitive d'informations, obtenu avec un noyau maxitif π, et les résultats des extractions sommatives d'informations obtenus avec les noyaux sommatifs de la famille $\mathcal{M}(\pi)$. Les résultats de ces deux types d'extraction d'informations sont obtenus respectivement par l'intégrale de Choquet et l'espérance mathématique.

Dans cette section, nous nous intéressons donc aux liens entre ces opérateurs d'agrégation, que sont l'intégrale de Choquet et l'espérance mathématique.

II.4.1 Intégrale de Choquet et espérance mathématique

Nous avons opté pour l'intégrale de Choquet, parce que dans l'extraction maxitive d'informations, cet opérateur propage l'imprécis contenu dans le choix d'un noyau maxitif à l'information extraite Y.

L'intégrale de Choquet et l'espérance mathématique sont définies pour toutes les fonctions bornées de Ω. L'ensemble de ces fonctions bornées est noté $\mathcal{L}(\Omega)$. Les capacités de Choquet et les mesures de probabilité sont définies pour des évènements de $\mathcal{B}(\Omega)$ qui est un sous ensemble de $\mathcal{L}(\Omega)$.

On peut interpréter l'intégrale de Choquet (II.13) par rapport à la capacité ν, comme une extension de ν définie sur $\mathcal{B}(\Omega)$, à l'ensemble des fonctions de $\mathcal{L}(\Omega)$. De même, l'espérance mathématique $\mathbb{E}_\kappa$ par rapport à la mesure de probabilité P_κ peut être interprétée comme une extension de P_κ, définie sur $\mathcal{B}(\Omega)$, à l'ensemble des fonctions de $\mathcal{L}(\Omega)$.

Dans le cadre de la théorie de Peter Walley, l'extension naturelle capte cette idée. En effet dans cette théorie, où Ω est fini, l'extension naturelle consiste à prolonger une prévision inférieure $\underline{P}$, définie pour un ensemble de paris $\mathcal{K} \subseteq \mathcal{L}(\Omega)$, à une prévision inférieure cohérente $\underline{E}$ sur $\mathcal{L}(\Omega)$. L'extension naturelle d'une prévision supérieure $\overline{P}$ est notée $\overline{E}$.

Il est intéressant de noter que l'extension naturelle d'une mesure de probabilité, définie pour des évènements de $\mathcal{B}(\Omega) \subseteq \mathcal{L}(\Omega)$, est l'espérance mathématique pour l'ensemble des fonctions de $\mathcal{L}(\Omega)$:

Théorème II.11 (Théorème de l'espérance mathématique)
Soit P une mesure de probabilité définie sur $\mathcal{B}(\Omega)$ associée à une distribution de probabilité κ. Soit E son extension naturelle (définition I.18). On a :

$$\forall f \in \mathcal{L}(\Omega),\ E(f) = \mathbb{E}_\kappa(f).$$

Ce théorème ainsi que sa démonstration peuvent être trouvés à la section 3.2.2 p.129 du livre de Peter Walley [105].

Par ailleurs, l'extension naturelle aux fonctions de $\mathcal{L}(\Omega)$, au sens de Peter Walley (I.108), d'une capacité de Choquet, définie pour des évènements de $\mathcal{B}(\Omega) \subseteq \mathcal{L}(\Omega)$, est l'intégrale de Choquet :

Théorème II.12 (Théorème de l'intégrale de Choquet)
Soit ν une capacité (i.e. une mesure non-additive) définie sur $\mathcal{B}(\Omega)$. Soit C son extension naturelle (définition I.18). On a :

$$\forall f \in \mathcal{L}(\Omega),\ C(f) = (C)\int_{\Omega} f d\nu.$$

On peut retrouver ce théorème et sa démonstration dans [54], théorème 5.5.

Le corollaire de ce théorème qui va nous intéresser est celui qui s'applique aux mesures de possibilité et de nécessité :

Corollaire II.13 (Extension naturelle et noyau maxitif)
Soit π un noyau maxitif défini sur Ω discret. Soient Π_π et N_π ses mesures de possibilités et de nécessité associées. Soient $\underline{E}$ l'extension naturelle de N_π et $\overline{E}$ l'extension naturelle de Π_π. $\forall f \in \mathcal{L}(\Omega)$, on a :

$$\underline{E}(f) = \mathbb{C}^c_\pi(f) = (C)\int_{\Omega} f dN_\pi,$$
$$\overline{E}(f) = \mathbb{C}_\pi(f) = (C)\int_{\Omega} f d\Pi_\pi.$$

L'extension naturelle, au sens de Peter Walley ou de Bruno de Finetti, conserve l'imprécision originale de l'évaluation d'une prévision inférieure $\underline{P}$ (ou supérieures $\overline{P}$) lors de son extension à $\underline{E}$ (ou supérieures $\overline{E}$). Ce qui nous intéresse, c'est la propagation de l'imprécision originale faite sur une mesure de possibilité à son extension naturelle que réalise l'intégrale de Choquet.

Par quelle mécanique cette imprécision est elle propagée ? En fait, la famille d'extensions naturelles des prévisions linéaires notée $E(\mathcal{M}(\underline{P}))$ est la même que la famille des prévisions linéaires dominées par $\underline{E}$, qui est $\mathcal{M}(\underline{E})$, i.e. $E(\mathcal{M}(\underline{P})) = \mathcal{M}(\underline{E})$. Pour notre méthode d'extraction maxitive d'informations, ce qui nous intéresse, c'est de voir que la famille des espérances mathématiques, que l'on peut noter $\mathbb{E}_{(\mathcal{M}(\pi^\theta))}$ obtenues à partir des noyaux sommatifs de $\mathcal{M}(\pi^\theta)$ est la même que la famille des éléments de l'intervalle $Y(\theta)$, que l'on peut également noter $\mathcal{M}(\mathbb{C}_{\pi^\theta})$, par similarité avec les notations de Peter Walley. Autrement dit, $\mathbb{E}_{(\mathcal{M}(\pi^\theta))} = \mathcal{M}(\mathbb{C}_{\pi^\theta})$.

Le théorème général [105] qui nous permet de conclure ce lien fort entre l'intégrale de Choquet et l'espérance mathématique est le suivant :

Théorème II.14 (Théorème de l'extension naturelle)
Soit $\underline{P}$ une prévision inférieure définie sur $\mathcal{K} \subseteq \mathcal{L}(\Omega)$, telle que $\mathcal{M}(\underline{P}) \neq \emptyset$. Soit $\underline{E}$ l'extension naturelle de $\underline{P}$ sur $\mathcal{L}(\Omega)$. Soit $\overline{E}$ l'extension naturelle de $\overline{P}$ (le conjugué de $\underline{P}$) sur $\mathcal{L}(\Omega)$. On note E_P, l'extension naturelle d'une prévision linéaire P qui fait partie de la famille $\mathcal{M}(\underline{P})$. $\forall f \in \mathcal{L}(\Omega)$, on a :

$$\underline{E}(f) = \min\{E_P(f) : P \in \mathcal{M}(\underline{P})\},$$
$$\overline{E}(f) = \max\{E_P(f) : P \in \mathcal{M}(\underline{P})\}.$$

Un corollaire du théorème II.14 va particulièrement nous intéresser dans le cadre de l'extraction maxitive d'informations :

Corollaire II.15 (Théorème discret de propagation de l'imprécis)
Soit π un noyau maxitif discret sur Ω, $\forall f \in \mathcal{L}(\Omega)$, on a :

$$\mathbb{C}_\pi^c(f) = \min\{\mathbb{E}_\kappa(f) : \kappa \in \mathcal{M}(\pi)\},$$
$$\mathbb{C}_\pi(f) = \max\{\mathbb{E}_\kappa(f) : \kappa \in \mathcal{M}(\pi)\}.$$

Ce théorème nous montre que, pour le cas où Ω est discret, le résultat de l'extraction maxitive d'informations avec un noyau maxitif contient tous les résultats des extractions sommatives avec les noyaux de la famille $\mathcal{M}(\pi)$ (cf. expression (II.19)) équivalente à π.

Dans le cas continu, ce théorème existe également. Il est dû à David Schmeidler et Dieter Denneberg qui ont prouvé le théorème II.16 ([87] proposition 3 et [21] proposition 10.3) pour des capacités concaves.

Théorème II.16 (Théorème de Schmeidler-Denneberg)
La capacité ν est concave si et seulement si, $\forall f \in \mathcal{L}(\Omega)$, on a :

$$(C)\int_\Omega f d\nu^c = \min\{\mathbb{E}_\kappa(f) : P_\kappa \in \mathcal{M}(\nu)\},$$
$$(C)\int_\Omega f d\nu = \max\{\mathbb{E}_\kappa(f) : P_\kappa \in \mathcal{M}(\nu)\}.$$

Un noyau maxitif définissant une mesure de possibilité, qui est un cas particulier de capacité concave, le corollaire au théorème de Schmeidler-Denneberg qui va nous intéresser est le suivant :

Corollaire II.17 (Théorème continu de propagation de l'imprécis)
Soit π un noyau maxitif continu sur Ω, $\forall f \in \mathcal{L}(\Omega)$, on a :

$$\mathbb{C}_\pi^c(f) = \min\{\mathbb{E}_\kappa(f) : \kappa \in \mathcal{M}(\pi)\},$$
$$\mathbb{C}_\pi(f) = \max\{\mathbb{E}_\kappa(f) : \kappa \in \mathcal{M}(\pi)\}.$$

Ce théorème nous montre que pour le cas où Ω est continu, le résultat de l'extraction maxitive d'informations avec un noyau maxitif contient tous les résultats des extractions sommatives avec les noyaux de la famille $\mathcal{M}(\pi)$ (cf. expression (II.19)) équivalente à π.

Les corollaires peuvent se résumer par le théorème suivant :

Théorème II.18 (Théorème fondamental de l'extraction maxitive d'informations)

Soit x une fonction de données. Soit π^θ un noyau maxitif qui régit l'extraction maxitive d'informations en θ, dont le résultat est noté $Y(\theta)$. Pour un noyau sommatif κ de $\mathcal{M}(\pi^\theta)$, on notera $y^\kappa(\theta)$ le résultat de l'extraction sommative avec κ. Les corollaires II.15 et II.17 nous montrent que $\forall \kappa \in \mathcal{M}(\pi^\theta)$, on a :

$$y^\kappa(\theta) \in Y(\theta). \tag{II.37}$$

II.4.2 Extraction maxitive d'informations et analyse bayésienne robuste

L'analyse bayésienne [3, 9] a pour objet principal d'estimer des paramètres β, au travers d'une méthode d'affinement de l'état de connaissance dans lequel on se trouve sur ces paramètres. En fait, ces méthodes constituent essentiellement un modèle d'apprentissage qu'on peut décrire en trois étapes :

1. un état initial d'incertitude sur les paramètres, représenté par une loi probabiliste dite *a priori* sur ces paramètres, noté $p(\beta)$,
2. une récolte de nouvelles données x et d'informations sur la façon dont vont interagir les données avec les paramètres. Ces informations sont résumées par la donnée de la loi de probabilité des données, lorsque l'on connaît les paramètres, i.e. par $p(x|\beta)$,
3. un état d'incertitude révisé (a posteriori), probabiliste lui aussi, sur les paramètres, obtenu par :

$$p(\beta|x) = \frac{p(x|\beta)p(\beta)}{\int p(x|\beta)p(\beta)d\beta}.$$

L'analyse bayésienne robuste tend à étudier l'impact, la sensibilité de la méconnaissance des a priori $p(\beta)$ et $p(x|\beta)$ sur les résultats des estimateurs bayésiens. Leur solution : utiliser des familles de distributions de probabilités a priori au lieu d'une seule distribution. Ces familles sont plus en adéquation avec les savoirs experts qui sont à l'origine de la définition des distributions a priori. L'école bayésienne est une école subjectiviste. Nous renvoyons à [5, 6, 4, 39] pour des exemples et discussions sur les familles de distributions a priori.

En extraction sommative d'information d'information, on fait l'hypothèse que les données sont transformées en informations par l'intermédiaire d'un noyau sommatif. Dans le but d'étudier l'impact de la méconnaissance de ce noyau sommatif, nous proposons, dans l'approche maxitive, de remplacer le noyau sommatif unique par une famille de noyaux sommatifs qu'est le noyau maxitif. L'analyse bayésienne robuste et l'extraction maxitive d'informations font le même effort de prise en compte des méconnaissances sur les hypothèses.

Si l'on revient à la particularisation de l'approche sommative aux statistiques, que nous avons présenté à la section I.4, il est intéressant de constater que nous avons dégagé seulement des applications de statistiques dites non-paramétriques. Dans les statistiques non-paramétriques, les méthodes ne font pas l'hypothèse que la loi sous-jacente aux données est une loi paramétrée (par exemple une gaussienne).

Ces méthodes sont classées dans la catégorie des statistiques robustes. Pourtant comme nous l'avons remarqué, les méthodes maxitives se rapprochent plus de l'idée de robustesse telle qu'elle est présentée dans l'analyse bayésienne robuste, que les méthodes sommatives. Elles font appel à une famille de distributions de probabilités plutôt qu'une unique distribution de probabilité. C'est pourquoi, l'approche maxitive alternative à ces méthodes non-paramétriques nous paraît plus qualifiable de robustes que les méthodes non-paramétriques, telles que l'estimateur de Parzen Rosenblatt.

II.5 Un indice de comportement des noyaux maxitifs en extraction maxitive d'informations : la granularité

Nous avons parlé de spécificité d'un noyau sommatif pour parler de son comportement. Cette même spécificité, évoquée en section II.3.2, est un marqueur naturel du comportement d'un noyau maxitif au sens où comportement et précisions sont confondus dans le cadre d'extraction maxitive d'informations.

Nous proposons ici une quantification du comportement des noyaux en extraction d'information par le biais d'un indice appelé granularité. Cet indice est un marqueur de

non spécificité du noyau.

Le terme "granularité", que nous proposons, s'inspire des travaux de Pawlak [75, 76] sur la granularité des rough sets et de Zadeh [118] sur la granulation de sous-ensembles flous.

La granularité d'un noyau maxitif π est définie par :

$$\Gamma(\pi) = \int_{\Omega} \pi(\omega) d\omega, \text{ dans le cas continu,} \tag{II.38}$$

$$\Gamma(\pi) = \sum_{i \in \Omega} \pi_i, \text{ dans le cas discret.} \tag{II.39}$$

Il est à noter que la granularité d'un noyau maxitif est mathématiquement identique à la cardinalité [30, 114] d'un sous-ensemble flou. Cette grandeur va mesurer à quel point les poids du noyau maxitif sont "éparpillés", autrement dit, à quel point le noyau maxitif est peu concentré, peu spécifique.

Nous avons justifié, en section II.3.2, que la spécificité d'un noyau maxitif caractérise la précision de ce noyau maxitif. De plus, l'imprécision d'un noyau maxitif est propagée à l'extraction maxitive d'informations (cf. théorème II.18). Voyons maintenant en quoi la spécificité va également caractériser la précision de l'extraction maxitive d'informations. Notons en préliminaire une propriété de l'intégrale de Choquet qui nous sera utile et que l'on peut retrouver dans [41, 21] :

Propriété II.19. *Soient μ et ν, deux capacités sur Ω. Soit f une fonction bornée sur Ω. On a*

$$\mu \leq \nu \Rightarrow (C) \int_{\Omega} f d\mu \leq (C) \int_{\Omega} f d\nu. \tag{II.40}$$

Examinons maintenant l'extraction maxitive locale d'informations en θ, à partir de données x, lorsque celle ci est effectuée avec deux noyaux maxitifs π_1 et π_2, tels que π_1 est plus spécifique que π_2, i.e. $\pi_1 \succeq_{SP} \pi_2$. On notera alors $Y_1(\theta) = [\mathbb{C}^c_{\pi_1}(x), \mathbb{C}_{\pi_1}(x)]$ et $Y_2(\theta) = [\mathbb{C}^c_{\pi_2}(x), \mathbb{C}_{\pi_2}(x)]$.

En associant les expressions (II.22) et (II.40), on arrive directement à la conclusion suivante :

$$\pi_1 \succeq_{SP} \pi_2 \Rightarrow \mathbb{C}^c_{\pi_2}(x) \leq \mathbb{C}^c_{\pi_1}(x) \leq \mathbb{C}_{\pi_1}(x) \leq \mathbb{C}_{\pi_2}(x). \tag{II.41}$$

Donc, lorsque π_1 est plus spécifique que π_2, le résultat de l'extraction maxitive d'informations obtenu avec π_1 est plus spécifique que celui obtenu avec π_2. On a même, d'après la relation (II.41), que lorsque π_1 est plus spécifique que π_2, $Y_1(\theta) \subseteq Y_2(\theta)$.

La spécificité d'un noyau maxitif caractérise donc non seulement son aptitude à se concentrer autour d'un point unique de son domaine de définition, mais caractérise aussi son comportement en extraction maxitive d'informations.

La granularité (II.38),(II.39) est un indice de non spécificité du noyau maxitif. Donc la granularité d'un noyau maxitif est un indice d'imprécision du noyau maxitif et de son extraction maxitive d'informations associée.

Chapitre III

Les noyaux maxitifs en traitement des données

Introduction

Dans le chapitre I, nous avons montré qu'un ensemble de méthodes de traitement du signal et de données statistiques peuvent être regroupées sous un formalisme unique que nous appelons **extraction sommative d'informations**. Ce formalisme fait intervenir une fonction positive qui est un cas particulier de distribution de probabilité : le **noyau sommatif**. Du choix du noyau sommatif dépend le fonctionnement et les résultats du traitement envisagé. Nous avons, au chapitre II, proposé une alternative à l'approche par noyau sommatif mettant en œuvre un autre type de noyau : le **noyau maxitif**, qui est un cas particulier de distribution de possibilité. L'ensemble des méthodes utilisant un tel noyau portent le nom d'**extraction maxitive d'informations**. L'intérêt de passer d'une approche sommative à une approche maxitive dépend de l'application. Nous proposons dans ce troisième chapitre de présenter des applications de la méthode d'extraction maxitive d'informations et de montrer, pour chacune d'elles, l'intérêt de passer de l'approche sommative à l'approche maxitive.

Nous présentons, en section III.1, la généralisation maxitive de la modélisation de l'acquisition, ainsi qu'une discussion sur une modélisation de l'erreur de mesure par noyau maxitif. Nous présentons, en section III.2, l'échantillonnage et la reconstruction d'un signal par noyaux maxitifs. Cette présentation est accompagnée d'une discussion sur l'intérêt et le but d'une éventuelle mesure maxitive. Nous présentons ensuite, en section III.3, une généralisation du filtrage d'un signal. Le résultat de ce filtrage maxitif imprécis est une disjonction de résultats de filtrages précis. La dérivation imprécise d'un signal y est proposée comme cas particulier.

Nous consacrons la section III.4 à des adaptations maxitives de méthodes et d'algorithmes de traitement d'images. Nous présentons une méthode de quantification du niveau de bruit sur une image basée sur le filtrage imprécis de l'image. Un lien mathématique est ensuite tendu entre une extension floue de la morphologie [11] et le filtrage imprécis. On prouve que le filtrage maxitif d'une image est équivalent à cette extension floue de la morphologie d'une image. Nous présentons ensuite une méthode de transformation rigide maxitive d'image, qui se base sur les passages Continu/Discret par noyaux maxitifs. Le résultat de cette transformation est l'ensemble des images transformées en accord avec les noyaux sommatifs que les noyaux maxitifs de passages Continu/Discret

représentent. Cette section se termine par une application de détection de contour, s'appuyant sur l'extraction maxitive d'informations, dont la robustesse vis-à-vis du bruit est très nettement améliorée par rapport aux approches sommatives usuelles (Canny-Deriche ou Shen-Castan).

En section III.5, nous présentons des généralisations maxitives des estimateurs à noyau de Parzen Rosenblatt de densité de probabilité et de fonction de répartition.

Nous terminons ce chapitre en proposant un indice du comportement d'un noyau sommatif en extraction sommative d'informations. C'est un indice de spécificité du noyau maxitif qui se base sur la transformation objective (cf. section II.3.4.3.1) d'un noyau sommatif en noyau maxitif.

III.1 Modélisation maxitive de l'acquisition et de l'erreur de mesure

III.1.1 Modéliser la mesure ou modéliser son inversion ?

Les données d'entrée d'un processus de traitement du signal proviennent la plupart du temps de systèmes de mesure appelés communément des capteurs. L'objet même d'un capteur est d'utiliser un dispositif de conversion d'énergie pour mesurer une grandeur à partir d'une autre. Pour donner un exemple simple : une mesure de déformation peut être obtenue facilement en utilisant un système comportant un cristal piezzo-électrique convertissant une contrainte en tension électrique. Lorsque la relation entre la grandeur à mesurer ou mesurande (dans cet exemple la déformation) et la sortie du dispositif de conversion ou mesure (dans cet exemple la tension) est linéaire (ou peut être supposée linéaire) alors la relation entre l'entrée et la sortie du processus peut être facilement modélisée par la réponse impulsionnelle du système. Il est alors possible d'utiliser la mesure comme si elle était la grandeur à mesurer en inversant simplement (on parle de déconvolution) la relation entrée-sortie. De fait, c'est le modèle inverse qui est généralement utilisé et non le modèle direct.

Le modèle sommatif permet de réaliser plutôt simplement cette inversion. Quand on connaît le passage du mesurande à la mesure, par un noyau sommatif κ^ω, l'inversion de ce passage consiste à retrouver les noyaux sommatifs $(\eta^u)_{u\in\Omega}$ utilisés pour passer d'une mesure en $\omega \in \Omega$ au mesurandes sur tout $u \in \Omega$ en résolvant ce qu'on appelle une équation orthogonale généralisée.

Les équations orthogonales généralisées s'inspirent des conditions d'orthogonalité des noyaux sommatifs utilisées pour l'adéquation parfaite (I.46) et (I.47). Dans ce type d'équation orthogonale généralisée, on cherche la famille de noyaux sommatifs $(\eta^u)_{u\in\Omega}$ qui, pour un noyau sommatif κ^ω donné, permet de vérifier la condition d'orthogonalité généralisée :

$$\int_{\Omega^c} \eta^u(v)\kappa^\omega(u)du = \delta^\omega(v), \tag{III.1}$$

où δ^ω est l'impulsion de Dirac en ω. Ce type d'équation est une généralisation des équations de convolution dont on peut trouver des méthodes de résolutions numériques simples dans [42]. Ces méthodes suivent en général le même schéma simple qui consiste en une résolution dans le domaine de Fourier de l'équation (III.1). Unser [102] propose notamment l'utilisation des B-splines pour faciliter cette résolution.

Il n'en reste pas moins que l'identification même de la réponse impulsionnelle modélisant le capteur, c'est-à-dire le passage du mesurande à la mesure, peut être sujette à débats. En effet, identifier une réponse impulsionnelle veut dire mettre en œuvre une série d'expérimentations nécessitant de

1. émettre une hypothèse sur une fonction paramétrée pouvant modéliser cette réponse,
2. identifier (généralement par régression) les paramètres de cette fonction.

On a donc à la fois une part d'arbitraire (choix de la fonction, choix de l'expérimentation) et une part d'aléatoire (bruit de mesure) qu'il est souvent difficile de répercuter sur le modèle. Pour continuer avec l'exemple précédent, une identification fréquentielle de la réponse impulsionnelle du capteur implique qu'on est capable d'exciter le système piezzo-électrique de façon pure avec des contraintes sinusoïdales dont les fréquences vont de 0 à l'infini, ainsi que de mesurer la réponse fréquentielle en tension.

III.1.2 Modélisation de la réponse impulsionnelle, l'approche maxitive

Un capteur est un dispositif permettant de mesurer, de façon généralement indirecte, une grandeur physique x, qui est le mesurande. Une "mesure" est la grandeur délivrée par le processus d'acquisition associé à ce capteur. Si le processus de mesure est linéaire et invariant par translation, il peut être modélisé par la convolution du signal entrant (associé ici au mesurande) avec la réponse impulsionnelle du capteur, c'est-à-dire par un processus d'extraction sommative d'informations (cf. section I.3.1). On peut alors, à une constante multiplicative près, associer la réponse impulsionnelle du capteur à un noyau sommatif κ^ω en ω. La mesure est alors modélisée par $\check{x}(\omega) = \mathbb{E}_{\kappa^\omega}(x)$.

La validité de ce modèle convolutif dépend de la possibilité d'identifier précisément la réponse impulsionnelle du capteur. Or comme nous l'avons dit, cette identification est sujette à caution. Dans ce contexte, l'utilisation d'un noyau maxitif en lieu et place d'un noyau sommatif permet de représenter une méconnaissance de la réponse impulsionnelle du capteur. La généralisation à l'approche maxitive est donnée par :

$$\check{X}(\omega) = [\underline{\check{X}}(\omega), \overline{\check{X}}(\omega)] = [\mathbb{C}^c_{\pi^\omega}(x), \mathbb{C}_{\pi^\omega}(x)]. \tag{III.2}$$

où π^ω est un noyau maxitif qui représente une famille de réponses impulsionnelles $\mathcal{M}(\pi^\omega)$.

III.1.3 Inversion du modèle maxitif de la mesure et erreur de mesure

Comme nous l'avons déjà dit, les capteurs actuels délivrent toujours des mesures précises. Pourtant, certains appareils de mesure (certains voltmètres par exemple) peuvent éventuellement afficher un intervalle, mais cet intervalle est issu d'une construction à partir d'une connaissance (d'un étalonnage) que le capteur a de ses défauts. Dans les approches probabilistes de traitement du signal, les erreurs de mesure sont directement modélisées par une distribution de probabilité. Cette modélisation de l'erreur vient se rajouter à la déconvolution du modèle sommatif de la mesure. Cette modélisation séparée des erreurs de mesure d'une part et de la mesure d'autre part produit une valeur imprécise du mesurande, qui correspond à la combinaison de la mesure et de l'erreur de mesure.

Définir l'inversion du modèle maxitif de mesure va nous permettre de proposer en un seul modèle, la combinaison de la mesure et de la connaissance des défauts du capteur. Ces défauts de capteurs pouvant représenter toutes les erreurs usuelles de mesures.

Le problème de l'inversion, si nous l'avons qualifié de simple en modélisation sommative de la mesure, s'avère plus compliqué avec le modèle maxitif. La mesure maxitive passe d'un mesurande précis à une mesure imprécise. Une inversion de type adéquation parfaite consisterait donc à déterminer l'opération qui permet de passer d'une mesure imprécise à un mesurande précis, quand on connait le modèle maxitif de la mesure. Cependant, la réalité nous impose un constat simple : les capteurs actuels délivrent toujours des mesures précises. Une inversion plus judicieuse de ce processus serait plutôt de passer d'une mesure précise à un mesurande imprécis, par l'intermédiaire d'une famille de modèles sommatifs inverses. On impacte ainsi, à partir d'une mesure précise, les défauts d'identification du modèle du capteur utilisé sur l'imprécision du mesurande estimé.

L'inversion du modèle maxitif de mesure que nous préconisons est une inversion un à un de tous les modèles sommatifs dominés par le modèle maxitif de la mesure. Des pistes de résolution de cette inversion constructive se trouvent dans le domaine de l'analyse convexe. En particulier, la notion d'Inf-convolution se rapproche singulièrement du calcul de la borne inférieure de l'extraction maxitive d'informations. Une étude approfondie des preuves d'existence de solutions d'équations d'Inf-convolution, que l'on peut trouver dans [55, 66], pourraient nous inspirer pour construire l'inverse d'une mesure maxitive. Une preuve d'existence par construction serait évidemment des plus plus utiles pour ce problème, mais ce sera pour un prochain épisode.

L'approche la plus simple est de ne pas passer par un noyau maxitif d'acquisition, mais directement de définir le noyau maxitif de passage des mesures à la réalité. On modélise ainsi, dans un seul objet mathématique, la mesure et ses erreurs.

III.2 Passages Continu/Discret par extraction maxitive d'informations

En traitement des données, et en traitement du signal particulièrement, l'échantillonnage et la reconstruction sont des étapes fondamentales pour la plupart des méthodes. Elles permettent des passages entre représentations continue et discrète des données. Ces étapes sont usuellement modélisées par l'extraction sommative d'informations. Nous présentons ici leur modélisation par l'extraction maxitive d'informations.

III.2.1 Reconstruction maxitive

Dans cette approche, le noyau maxitif discret utilisé π^ω, pour $\omega \in \Omega$, modélise une famille de noyaux sommatifs de reconstruction, noté $\mathcal{M}(\pi^\omega)$. Cette famille correspond à une méconnaissance sur le noyau de reconstruction à utiliser. Le résultat de la reconstruction maxitive d'un signal est un intervalle dont les bornes inférieure et supérieure correspondent aux valeurs inférieure et supérieure de l'ensemble des signaux reconstruits avec les noyaux sommatifs de la famille $\mathcal{M}(\pi^\omega)$ (cf. théorème II.18).

La modélisation sommative de la reconstruction d'un signal continu $\hat{x}$, par le noyau sommatif κ^ω, en $\omega \in \Omega$, à partir des données $x = (x_i)_{i=1,\dots,n}$, est donnée par $\hat{x}(\omega) =$

$\sum_{i=1}^{n} x_i \kappa_i^\omega = \mathbb{E}_{\kappa^\omega}(x)$. La modélisation maxitive est donnée, $\forall \omega \in \Omega$, par :

$$\hat{X}(\omega) = [\underline{\hat{X}}(\omega), \overline{\hat{X}}(\omega)] = [\mathbb{C}^c_{\pi^\omega}(x), \mathbb{C}_{\pi^\omega}(x)]. \tag{III.3}$$

où π^ω est le noyau maxitif qui représente une famille de noyaux sommatifs de reconstruction $\mathcal{M}(\pi^\omega)$.

III.2.2 Échantillonnage maxitif

Dans cette approche, le noyau maxitif continu utilisé π^k, pour $k \in \Omega = \{1, ..., p\}$, correspond à une famille de noyaux sommatifs d'échantillonnage, noté $\mathcal{M}(\pi^k)$. Cette famille correspond à une méconnaissance sur le noyau modélisant le processus d'échantillonnage, c'est-à-dire celui associé à la réponse impulsionnelle de l'échantillonneur. Le résultat de l'échantillonnage maxitif d'un signal est un intervalle dont les bornes inférieure et supérieure correspondent aux valeurs inférieure et supérieure de l'ensemble des signaux échantillonnés par les noyaux sommatifs de la famille $\mathcal{M}(\pi^k)$ (cf. théorème II.18).

La modélisation sommative de l'échantillonnage d'un signal $\check{x} = (\check{x}_i)_{i=1,...,n}$, par le noyau sommatif κ^k, en $k \in \Omega$, à partir d'un signal continu x est donnée par $\check{x}_k = \int_\Omega x(v)\kappa^k(v)dv = \mathbb{E}_{\kappa^k}(x)$. La modélisation maxitive est donnée, $\forall k \in \Omega$, par :

$$\check{X}_k = [\underline{\check{X}}_k, \overline{\check{X}}_k] = [\mathbb{C}^c_{\pi^k}(x), \mathbb{C}_{\pi^k}(x)]. \tag{III.4}$$

où π^k est le noyau maxitif qui représente une famille de noyaux sommatifs de reconstruction $\mathcal{M}(\pi^k)$.

III.3 Filtrage imprécis de signal

III.3.1 Principe et définitions

Le principe du filtrage des signaux est exposé en section I.3.5. Il met en jeu un signal d'entrée x, défini sur Ω, un filtre linéaire de réponse impulsionnelle η, et un signal de sortie $\tilde{x}$ défini sur Ω, égal au produit de convolution de x par η. C'est pourquoi on donne le nom de noyau de convolution à la fonction η. On a montré que tout filtrage linéaire, qu'il soit analogique ou numérique, pouvait être exprimé à partir d'extractions sommatives d'informations.

Pour le filtrage numérique d'un signal (I.60), $\forall k \in \{1, ..., p\}$,

$$\tilde{x}_k = a^{k+}\mathbb{E}_{\kappa^{k+}}(x) - a^{k-}\mathbb{E}_{\kappa^{k-}}(x), \tag{III.5}$$

où κ^{k+} et κ^{k-} sont les noyaux sommatifs obtenus à partir du noyau de convolution η^k et qui forment sa décomposition linéaire (cf. démonstration du théorème I.9). a^{k+} et a^{k-} sont les coefficients de cette décomposition.

Pour le filtrage analogique d'un signal (I.59), $\forall \omega \in \Omega$,

$$\tilde{x}(\omega) = a^{\omega+}\mathbb{E}_{\kappa^{\omega+}}(x) - a^{\omega-}\mathbb{E}_{\kappa^{\omega-}}(x), \tag{III.6}$$

où $\kappa^{\omega+}$ et $\kappa^{\omega-}$ sont les noyaux sommatifs obtenus à partir du noyau de convolution η^ω et qui forment sa décomposition linéaire (cf. démonstration du théorème I.9). $a^{\omega+}$ et $a^{\omega-}$ sont les coefficients de cette décomposition.

Le principe du filtrage maxitif des signaux est de remplacer les noyaux sommatifs $\kappa^{\omega+}$ et $\kappa^{\omega-}$ (resp. κ^{k+} et κ^{k-} dans le cas du filtrage numérique) qui forment la décomposition du noyau de convolution η par des noyaux maxitifs $\pi^{\omega+}$ et $\pi^{\omega-}$ (resp. π^{k+} et π^{k-} dans le cas du filtrage maxitif numérique).

Le filtrage maxitif numérique d'un signal est obtenu, $\forall k \in \{1, ..., p\}$, par :

$$\tilde{X}_k = a^{k+}[\mathbb{C}^c_{\pi^{k+}}(x), \mathbb{C}_{\pi^{k+}}(x)] \ominus a^{k-}[\mathbb{C}^c_{\pi^{k-}}(x), \mathbb{C}_{\pi^{k-}}(x)], \tag{III.7}$$

où $\ominus$ est l'extension de la soustraction aux intervalles.

Les bornes de l'estimation sont donc, $\forall k \in \{1, ..., p\}$,

$$\underline{\tilde{X}}_k = a^{k+}\mathbb{C}^c_{\pi^{k+}}(x) - a^{k-}\mathbb{C}_{\pi^{k-}}(x), \tag{III.8}$$

$$\overline{\tilde{X}}_k = a^{k+}\mathbb{C}_{\pi^{k+}}(x) - a^{k-}\mathbb{C}^c_{\pi^{k-}}(x), \tag{III.9}$$

Le filtrage maxitif analogique d'un signal est obtenu, $\forall \omega \in \Omega$, par :

$$\tilde{X}(\omega) = a^{\omega+}[\mathbb{C}^c_{\pi^{\omega+}}(x), \mathbb{C}_{\pi^{\omega+}}(x)] \ominus a^{\omega-}[\mathbb{C}^c_{\pi^{\omega-}}(x), \mathbb{C}_{\pi^{\omega-}}(x)]. \tag{III.10}$$

Les bornes de l'estimation sont donc, $\forall \omega \in \Omega$,

$$\underline{\tilde{X}}(\omega) = a^{\omega+}\mathbb{C}^c_{\pi^{\omega+}}(x) - a^{\omega-}\mathbb{C}_{\pi^{\omega-}}(x), \tag{III.11}$$

$$\overline{\tilde{X}}(\omega) = a^{\omega+}\mathbb{C}_{\pi^{\omega+}}(x) - a^{\omega-}\mathbb{C}^c_{\pi^{\omega-}}(x), \tag{III.12}$$

Dans de nombreux cas, la réponse impulsionnelle est positive (I.67) (modélisation d'un capteur, filtres de lissage...). Dans ce contexte, la réponse impulsionnelle du filtre peut s'exprimer, à une constante multiplicative près, comme un noyau sommatif, à savoir $\eta = a\kappa$. La généralisation du filtrage linéaire positif devient très directe :

Le filtrage maxitif positif discret d'un signal est obtenu, $\forall k \in \{1, ..., p\}$, par :

$$\tilde{X}_k = a^k[\mathbb{C}^c_{\pi^k}(x), \mathbb{C}_{\pi^k}(x)]. \tag{III.13}$$

Le filtrage maxitif positif continu d'un signal est obtenu, $\forall \omega \in \Omega$, par :

$$\tilde{X}(\omega) = a^{\omega}[\mathbb{C}^c_{\pi^{\omega}}(x), \mathbb{C}_{\pi^{\omega}}(x)]. \tag{III.14}$$

III.3.2 Dérivation d'un signal

Nous avons vu en section I.3.6, que la dérivation numérique d'un signal par noyau sommatif (qui s'appuie les passages C/D) s'exprime (cf. expression (I.72)) par :

$$d\tilde{x}_k = a^-\mathbb{E}_{d\eta^{k-}}(x) - a^+\mathbb{E}_{d\eta^{k+}}(x).$$

Soient π^{k-} et π^{k+}, des noyaux maxitifs qui dominent les noyaux sommatifs $d\eta^{k-}$ et $d\eta^{k+}$, formant la combinaison linéaire de la réponse impulsionnelle du filtre de dérivation discret $d\eta^k$. Autrement dit, $d\eta^{k-} \in \mathcal{M}(\pi^{k-})$ et $d\eta^{k+} \in \mathcal{M}(\pi^{k+})$. La dérivation maxitive discrète d'un signal x est obtenue, pour tout $k \in \{1, ..., p\}$, par

$$D\tilde{X}_k = a^-[\mathbb{C}^c_{\pi^{k-}}(x), \mathbb{C}_{\pi^{k-}}(x)] \ominus a^+[\mathbb{C}^c_{\pi^{k+}}(x), \mathbb{C}_{\pi^{k+}}(x)]. \tag{III.15}$$

III.4 Traitement maxitif des images

III.4.1 Quantification du niveau de bruit

Quand on parle du bruit sur une image, on se réfère généralement à des variations aléatoires sur la luminance mesurée. Ces variations peuvent avoir différentes causes : bruit thermique, saturation du capteur, échantillonnage, bruit électronique... Ces variations statistiques sont observables si l'on dispose d'une grande quantité d'acquisition de la même image. Seulement, en règle générale, on n'a qu'une seule image et les variations aléatoires doivent être estimées sur cette image unique.

Les approches usuelles considèrent que le bruit est indépendant de l'image et que la distribution des variations aléatoires induites par ce bruit peut faire l'objet d'un pré-calibrage. Dans la majorité des cas, le bruit est considéré comme gaussien, lorsqu'il provient de phénomènes physiques ou uniforme lorsqu'il s'agit de modéliser la quantification. Le modèle le plus proche de la réalité de la mesure de luminance reste cependant le processus de Poisson. Des hypothèses sur sa répartition sur l'image sont également nécessaire pour caractériser le bruit : par exemple, on fait souvent l'hypothèse que le bruit est stationnaire [73] ou que le niveau de bruit dépend de la valeur de niveau de gris de l'image (le cas d'un bruit de Poisson [57, 44, 63]).

L'utilisation d'un pré-calibrage du bruit est souvent impossible, particulièrement lorqu'un même algorithme doit traiter des images provenant de sources différentes et inconnues. C'est pourquoi on lui préfère généralement un estimateur local (ou global) de bruit basé sur l'ergodicité locale (ou globale) de celui-ci.

Notre approche se base exclusivement sur l'hypothèse d'ergodicité locale. Un signal est dit ergodique quand ses moments statistiques sont égaux à ses moments spatiaux. Rappelons que le moment d'ordre 1 est la moyenne et le moment d'ordre 2 est la variance. Un signal localement ergodique a, dans des voisinages prédéfinis[1], des moments statistiques locaux et des variations spatiales locales égaux. Comme la plupart des approches utilisant l'hypothèse d'ergodicité locale, notre approche consiste en une étude spatiale locale de l'image, permettant d'en estimer les variations statistiques et donc le bruit. Ce bruit impulsionnel est référencé dans la littérature des géostatistiques sous le nom d'effet pépite, que nous reprenons ici. Nous comparons notre méthode locale d'estimation de l'effet pépite à d'autre méthodes locales consistant à estimer ces variations dans un voisinage défini par un noyau sommatif.

III.4.1.1 L'effet pépite

Sous le principe d'effet pépite, il y a l'idée, qu'en moyenne, les valeurs de pixels d'une image échantillonnée spatialement proches se comportent d'une manière plus similaire que des valeurs de pixels de l'image éloignés. Le principe d'effet pépite est issu des géostatistiques [64, 40]. L'hypothèse sous-jacente à cette approche, c'est que la variabilité spatiale d'un sous domaine de l'image, noté A, est supposée refléter conjointement la variabilité locale intrinsèque de l'image réelle continue sous-jacente à l'image échantillonnée ainsi que les erreurs de mesures. Typiquement, la variabilité spatiale augmente avec la taille du

[1] L'ergodicité locale est normalement définie pour une taille de voisinage. Dans ce manuscrit, quand nous évoquons l'ergodicité locale, la taille du voisinage de l'ergodicité locale correspond à la taille du noyau employé.

sous domaine A. Au contraire, la variabilité due aux erreurs de mesures ne devrait pas dépendre de la taille de A.

Notons A^k_Δ le voisinage du pixel k de rayon Δ. On note, de façon symbolique, $V(A^k_\Delta)$, la variabilité du signal dans le voisinage A^k_Δ. Sous l'hypothèse précédente, cette variabilité vérifie :

$$\lim_{\Delta \to 0} V(A^k_\Delta) = v_k. \tag{III.16}$$

Lorsqu'il est possible de supposer que v_k ne dépend pas de k, cette limite porte le nom d'effet pépite [40]. Que v_k dépende ou non de k, sa mesure directe n'est pas possible, car on ne peut pas calculer $V(A^k_\Delta)$ si Δ est inférieur au pas d'échantillonnage h.

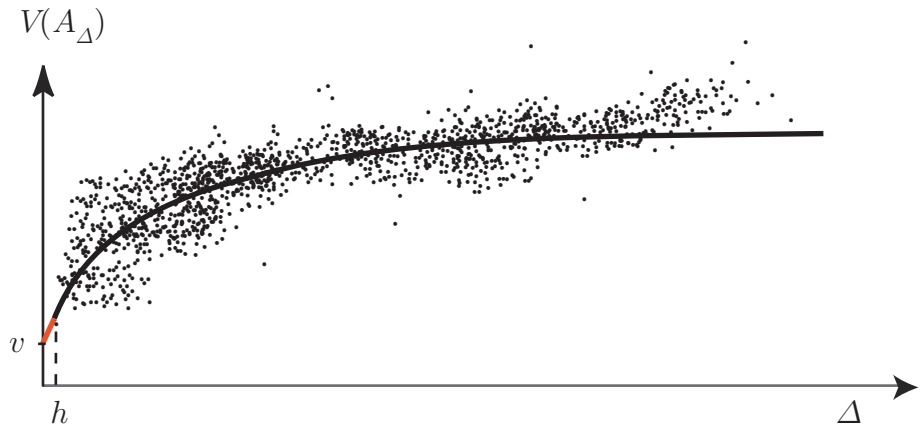


Fig. III.1 : Variogramme pour l'estimation de la variabilité v par effet pépite

Une des solutions les plus utilisées en géostatistiques, pour estimer l'effet pépite, exploite une représentation de l'évolution de la variabilité en fonction de la distance appelée variogramme. Un variogramme, tel que représenté en figure III.1, consiste à calculer, pour des distances $\Delta > h$, la moyenne $V(A_\Delta)$ des $V(A^k_\Delta)$ sur l'ensemble des pixels k de l'image. On résume alors l'information du variogramme en réalisant une régression sur l'ensemble des couples $(\Delta, V(A_\Delta))$. La courbe ainsi obtenue est alors prolongée jusqu'à $\Delta = 0$ pour approximer v, l'effet pépite.

Hors les géostatistiques, le variogramme est très peu utilisé en traitement classique des images car il présuppose une forme particulière aux variations (outre la stationnarité d'ordre 2), nécessitant le choix d'une fonction. Le choix d'un tel modèle n'est pas forcément justifié en traitement classique d'images, alors qu'il l'est en géostatistiques. Le résultat de l'estimation de v dépend beaucoup du choix de la fonction de régression. Cette limite particulièrement invalidante fait, qu'en traitement d'images, on lui préfère des méthodes plus locales.

III.4.1.2 Quantification par voisinage sommatif

Sous une hypothèse plus raisonnable d'ergodicité locale, il est possible d'estimer directement l'effet pépite en utilisant un voisinage pondéré sommatif invariant par translation κ (cf. expression (I.18)). La variabilité de l'image, au pixel k, peut être estimée par :

$$v_k = \sqrt{\sum_{i=1}^{n} (I_i - \hat{I}_k)^2 \kappa^k_i}, \tag{III.17}$$

si la variabilité est mesurée par la variance et

$$v_k = \sum_{i=1}^{n} |I_i - \hat{I}_k| \kappa_i^k, \tag{III.18}$$

si la variabilité est mesurée par l'erreur moyenne. $\hat{I}_k$, la moyenne locale pondérée de l'image, est obtenue par :

$$\hat{I}_k = \sum_{i=1}^{n} I_i \kappa_i^k = \mathbb{E}_{\kappa^k}(I). \tag{III.19}$$

On remarque que la moyenne pondérée est un cas particulier d'extraction sommative d'informations. Cette méthode se base sur l'idée que le rayon d'action du noyau sommatif est relativement petit par rapport à la distance d'échantillonnage. C'est-à-dire que le noyau sommatif employé agit sur un petit nombre de pixels autour du pixel considéré. Si les erreurs dues au bruit aléatoire sont supérieures aux effets conjoints de l'échantillonnage, de la quantification et des variations du signal, alors une petite taille relative du rayon d'action du noyau κ par rapport au pas d'échantillonnage h permet de considérer que les variations mesurées sont dues au bruit statistique de l'image.

III.4.1.3 Notre approche

Notre approche utilise un noyau maxitif π, invariant par translation, pour calculer, de façon imprécise, la moyenne de l'image au pixel k. La conjecture sur laquelle s'appuie notre approche est que l'imprécision d'estimation locale de la moyenne est plus corrélée avec les variabilités d'un signal dues au bruit que les mesures de variabilité classiques que sont la variance (III.17) ou l'écart type (III.18).

Comme pour les méthodes précédentes, est exploitée l'idée d'un voisinage pondéré dont le rayon d'action est suffisamment petit par rapport au pas d'échantillonnage pour pouvoir considérer qu'on observe une variation due au bruit et non au signal. L'estimation imprécise de la moyenne est donnée par :

$$[\underline{\hat{I}}_k, \overline{\hat{I}}_k] = [\mathbb{C}^c_{\pi^k}(I), \mathbb{C}_{\pi^k}(I)], \tag{III.20}$$

et on estime le niveau de bruit par :

$$\lambda_k = \mathbb{C}_{\pi^k}(I) - \mathbb{C}^c_{\pi^k}(I). \tag{III.21}$$

Nous avons lié, au chapitre II, l'imprécision du résultat de l'extraction maxitive d'informations à la spécificité du noyau maxitif employé. Ce lien n'est pas exclusif, au sens où d'autres facteurs entrent en ligne de compte pour déterminer l'imprécision d'estimation λ_k (III.21). Cette remarque est essentielle pour justifier notre approche. L'idée est la suivante. Supposons que le voisinage défini autour du pixel k par le noyau maxitif π^k soit suffisamment spécifique comparé à la résolution du signal, pour que localement, le signal puisse être considéré comme constant. Alors toutes les moyennes du signal calculées avec les noyaux sommatifs dominés par π^k seraient égales, et on aurait donc $\lambda_k = 0$. Dans le sens inverse, si $\lambda_k \neq 0$, cela signifie qu'il y a, suivant les noyaux sommatifs dominés par π^k, des désaccords sur la moyenne du signal. Plus il y a de variations locales du signal, plus il y a de désaccord sur la valeur moyenne du signal. La grandeur de λ_k est donc, à

spécificité de π^k fixée, un marqueur des variations locales du signal. C'est d'ailleurs pour cela que l'on parle d'estimation de niveau de bruit plutôt que d'estimation du bruit. On constatera d'ailleurs, dans l'expérience III.4.1.4, que les λ^k obtenus n'ont pas n'ont pas le même ordre de grandeur que les variations statistiques de l'image, mais y sont fortement corrélés.

Justifions maintenant notre conjecture par l'expérience.

III.4.1.4 Expériences

III.4.1.4.1 Matériel

Pour cette expérience, nous avons rempli un fantôme de cerveau de Hoffman (Data Spectrum Corporation) avec une solution de 99m technetium (148MBq/L). Le fantôme est placé devant l'un des deux détecteurs d'une gamma camera à double tête avec collimateur à orifices parallèles (INFINIA, General Electric Healthcare). Nous avons effectué 1000 acquisitions (temps d'acquisition : 1 seconde ; comptages moyens par image : 1500 comptages, images 128×128 pour satisfaire les conditions de Shannon), représentant 1000 mesures d'une image aléatoire $2D$ dont le bruit est généralement supposé suivre une loi de Poisson.

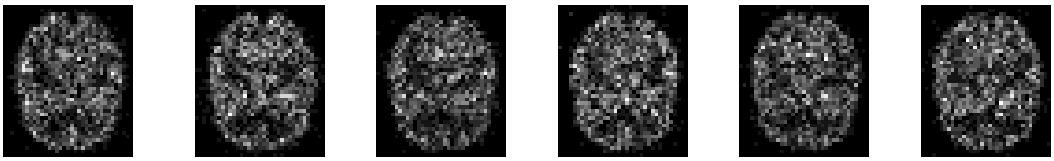


Fig. III.2 : Six images parmi les 1000 acquisitions du fantôme de cerveau de Hoffman

Le temps d'acquisition étant très court, les images sont très bruitées. On note $I_{k,j}$ la valeur mesurée de l'activité sur le k-ième pixel dans la j-ième acquisition de l'image. La figure III.2 ne présente que les partie centrale 40×35 des images acquises du fantôme de Hoffman.

III.4.1.4.2 Expériences

Dans cette expérience, nous tentons de montrer la capacité de notre approche basée sur les noyaux maxitifs, présentée en section III.4.1.3, à quantifier les variations statistiques dues au bruit.

Dû au fait que la radioactivité est décrite par une distribution de Poisson, on ne peut pas considérer que le bruit soit stationnaire sur toute l'image. Le rapport signal sur bruit étant très faible, la variation locale du niveau de radioactivité dans un voisinage de pixel reste très corrélée avec les variations statistiques dues au bruit d'acquisition.

D'une part, on peut estimer les variations statistiques de l'activité au pixel k par sa variance statistique empirique σ_k suivant les 1000 acquisitions :

$$\sigma_k = \sqrt{\frac{1}{999}\sum_{j=1}^{1000}(I_{k,j} - m_k)^2}, \tag{III.22}$$

où m_k est la moyenne statistique au k-ème pixel :

$$m_k = \frac{1}{1000} \sum_{j=1}^{1000} I_{k,j}. \tag{III.23}$$

D'autre part, la variation locale de mesure dans le voisinage du pixel k de la j-ième image peut être estimée par l'expression (III.17) avec un noyau sommatif très spécifique. La même expérience réalisée avec l'expression (III.18) a donné des résultats similaires. On propose, dans cette expérience, d'estimer deux fois cette variation spatiale locale, avec deux noyaux sommatifs différents : $\gamma_{k,j}$ est obtenu à partir du noyau sommatif uniforme 3×3 ; $\delta_{k,j}$ est obtenu à partir d'un noyau gaussien de variance égale à 1.6. Ce choix de variance pour ce second noyau est adapté au noyau uniforme par la méthode d'adaptation par granularité présentée dans [62]. Rappelons que cette méthode est plus à même d'estimer le bruit local que la méthode d'effet pépite présentée à la section précédente.

En parallèle, on calcule, pour chaque image j, les valeurs moyennes imprécises locales de l'image $[\hat{\underline{I}}_{k,j}, \hat{\overline{I}}_{k,j}]$ dans le voisinage du pixel k, défini par le noyau maxitif π^k, obtenues par l'expression (III.20). Les variations locales au pixel k de l'image j sont ainsi estimées par les longueurs $\lambda_{k,j}$ de ces moyennes imprécises :

$$\lambda_{k,j} = \hat{\overline{I}}_{k,j} - \hat{\underline{I}}_{k,j}. \tag{III.24}$$

Le noyau π^k utilisé est le noyau translaté en k du noyau maxitif suivant :

$$\pi_{3\times 3} = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{pmatrix} \tag{III.25}$$

Le noyau maxitif $\pi_{3\times 3}$ est obtenu par prod-séparabilité (cf. définition II.6), à partir des deux mêmes noyaux maxitifs π définis sur les axes des abscisses et des ordonnées de l'image I, par :

$$\pi = \begin{pmatrix} 0.5 & 1 & 0.5 \end{pmatrix} \tag{III.26}$$

Le noyau maxitif π domine, entre autre, tous les voisinages pondérés sommatifs symétriques centrés de valeur non nulle sur trois pixels, issus de l'échantillonnage d'un noyau sommatif continu monomodal, symétrique et de support trois pixels.

Notre but, avec cette expérimentation, est de tester si la distribution des écarts-types statistiques σ_k est corrélée avec les distributions des $\gamma_{k,j}$, $\delta_{k,j}$ et $\lambda_{k,j}$. Nous effectuons cette comparaison sur les valeurs brutes, mais aussi sur les valeurs moyennes, pour chaque pixel k, des distributions des mesures de variations $\gamma_{k,j}$, $\delta_{k,j}$ et $\lambda_{k,j}$, obtenues comme suit :

$$\tilde{\gamma}_k = \frac{1}{1000} \sum_{j=1}^{1000} \gamma_{k,j},$$

$$\tilde{\delta}_k = \frac{1}{1000} \sum_{j=1}^{1000} \delta_{k,j},$$

$$\tilde{\lambda}_k = \frac{1}{1000} \sum_{j=1}^{1000} \lambda_{k,j}.$$

La figure III.3 représente le nuage de points $(\tilde{\gamma}_k, \sigma_k)$, la figure III.4 représente le nuage de points $(\tilde{\delta}_k, \sigma_k)$ et la figure III.5 représente le nuage de points $(\tilde{\lambda}_k, \sigma_k)$. Nous avons choisi de ne pas représenter les nuages de points associés aux $\gamma_{k,j}$, $\delta_{k,j}$ et $\lambda_{k,j}$ car ces figures sont très difficilement interprétables visuellement.

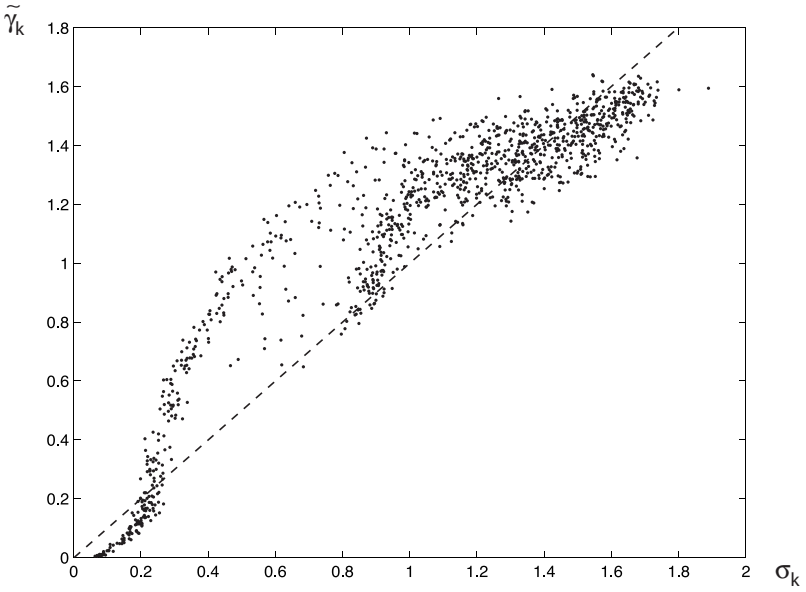


Fig. III.3 : Nuage de points représentant les variations locales $\tilde{\gamma}_k$ en fonction des variations statistiques σ_k pour tous les pixels k

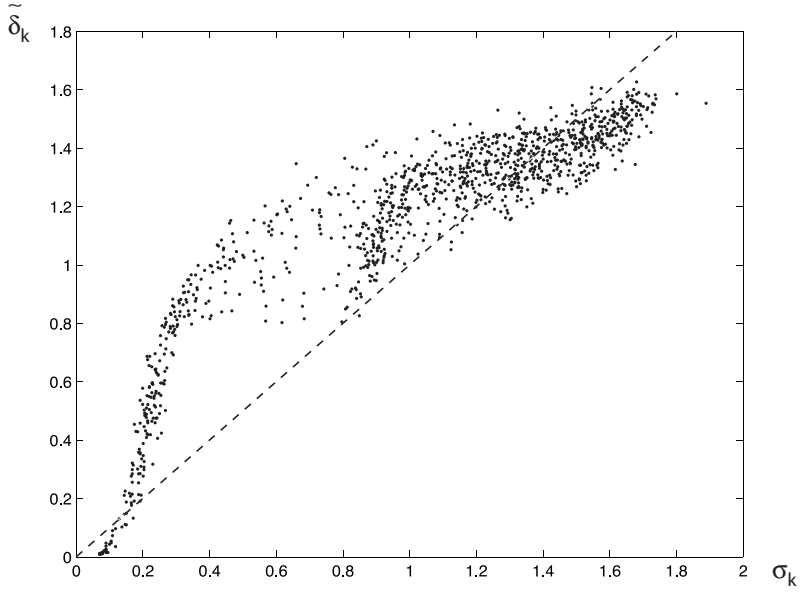


Fig. III.4 : Nuage de points représentant les variations locales $\tilde{\delta}_k$ en fonction des variations statistiques σ_k pour tous les pixels k

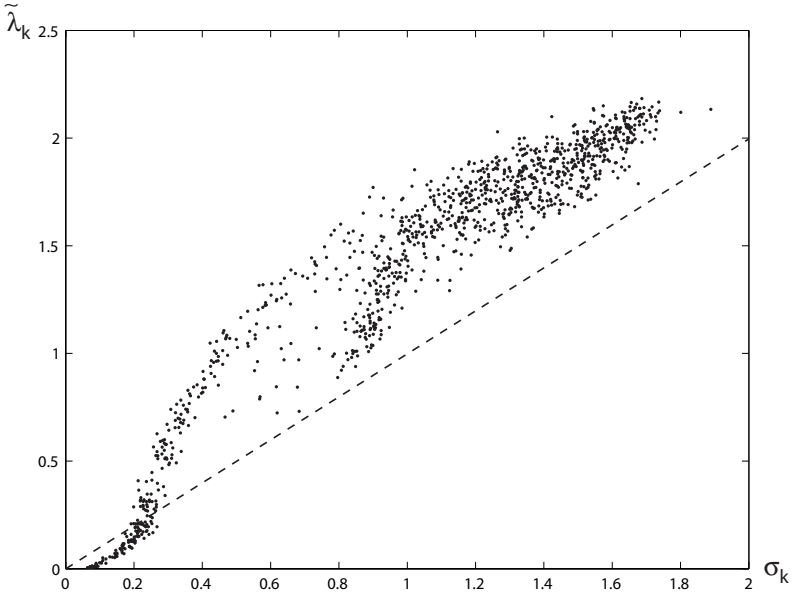


Fig. III.5 : Nuage de points représentant les variations locales $\tilde{\lambda}_k$ en fonction des variations statistiques σ_k pour tous les pixels k

Les figures III.3, III.4 et III.5 montrent nettement une corrélation, en moyenne, entre ces estimations de variation locale et les écarts-types statistiques σ_k.

Sur chacune des figures, nous avons rajouté la droite identité. Cette droite est utile pour comparer les ordres de grandeur des variations statistiques σ_k et des variations locales estimées. Elle permet également d'observer les zones de sous-estimation et de sur-estimation.

Le choix de la valeur 1.6, pour la variance du noyau sommatif gaussien, est appropriée à cette estimation, car les valeurs obtenues par cet estimateur $\tilde{\gamma}_k$ en figure III.4 sont du même ordre de grandeur que les variations statistiques σ_k. La même remarque peut être faite pour le choix du noyau uniforme 3×3 avec les estimateurs $\tilde{\gamma}_k$ en figure III.3. Sur la figure III.5, on constate que $\tilde{\lambda}_k$ sur-estime légèrement les variations statistiques σ_k.

Afin de comparer objectivement les trois mesures de variations $\gamma_{k,j}$, $\delta_{k,j}$ et $\lambda_{k,j}$, on calcule les coefficients de corrélation de Pearson, Spearman et Kendall de ces mesures par rapport à la distribution des σ_k. Le même calcul est effectué pour les mesures moyennes $\tilde{\gamma}_k$, $\tilde{\delta}_k$ et $\tilde{\lambda}_k$.

Comme on peut le constater sur le tableau III.1, les trois mesures $\tilde{\gamma}_k$, $\tilde{\delta}_k$ et $\tilde{\lambda}_k$ sont très corrélées avec σ_k. Comme on pouvait s'y attendre, la corrélation entre σ_k et les mesures $\gamma_{k,j}$, $\delta_{k,j}$ et $\lambda_{k,j}$ est plus faible que pour leurs moyennes statistiques. Ces valeurs sont cependant suffisantes pour montrer la dépendance entre ces mesures de variations spatiales locales et les variations statistiques du signal. On constate que $\lambda_{k,j}$ est plus corrélée avec σ_k que ne le sont $\gamma_{k,j}$ et $\delta_{k,j}$. On peut également constater que $\tilde{\lambda}_k$ apparaît toujours plus corrélée avec σ_k que ne le sont $\tilde{\gamma}_k$ et $\tilde{\delta}_k$. On peut donc conclure que la mesure $\lambda_{k,j}$, basée sur l'extraction maxitive d'informations, est un meilleur quantificateur de niveau de bruit, pour cette expérience, que les mesures de type sommatif $\gamma_{k,j}$ et $\delta_{k,j}$.

	$\gamma_{k,j}$	$\tilde{\gamma}_k$	$\delta_{k,j}$	$\tilde{\delta}_k$	$\lambda_{k,j}$	$\tilde{\lambda}_k$
Pearson	0.70	0.93	0.64	0.90	0.71	0.96
Spearman	0.64	0.92	0.63	0.90	0.67	0.95
Kendall	0.47	0.77	0.47	0.75	0.51	0.81

Tab. III.1 : Coefficients de corrélation entre la variabilité statistique et les mesures de variabilité locales proposées.

III.4.2 Filtrage maxitif et morphologie floue d'une image

III.4.2.1 Filtrage maxitif d'une image

Considérons le filtrage discret d'une image numérique, dont on a les mesures discrètes $(I_i)_{i=1,...,n}$, en n positions régulièrement réparties. On suppose en fait ici qu'est effectué un filtrage de l'image avec un noyau de convolution κ sommatif discret. Le filtrage maxitif de cette image est réalisé par l'utilisation d'un noyau maxitif π à la place du noyau sommatif κ. On obtient les images filtrées inférieures et supérieures suivantes, pour tout $k \in \{1, ..., n\}$,

$$\underline{\tilde{I}}_k = \sum_{i=1}^{n} (I_{\sigma(i)} - I_{\sigma(i-1)}) N_{\pi^k}(A_i), \tag{III.27}$$

$$\overline{\tilde{I}}_k = \sum_{i=1}^{n} (I_{\sigma(i)} - I_{\sigma(i-1)}) \Pi_{\pi^k}(A_i), \tag{III.28}$$

avec $A_i := \{\sigma(i), ..., \sigma(n)\}$, $\sigma(0) = 0$ et σ la permutation sur Ω, telle que $I_{\sigma(1)} \leq ... \leq I_{\sigma(n)}$. Les A_i sont des coalitions de pixels de l'image basées sur l'ordre σ, et par convention, $I_0 = 0$.

L'image intervalliste filtrée obtenue correspond à l'ensemble des images que l'on aurait obtenues avec des filtres de réponses impulsionnelles $\kappa \in \mathcal{M}(\pi)$.

III.4.2.2 Morphologie floue sur une image

Les opérations de base de la morphologie mathématique sont deux opérations duales appelées respectivement érosion et dilatation. Ces noms proviennent de l'effet produit sur une image binaire. Des combinaisons de ces opérateurs permettent de filtrer les images du point de vue morphologique, c'est-à-dire supprimer (ou rehausser) des formes sur l'image comparables aux éléments structurants utilisés dans les opérations morphologiques d'érosion et de dilatation. C'est le cas de l'ouverture (succession d'une érosion et d'une dilatation utilisant le même élément structurant) et de la fermeture (succession d'une dilatation et d'une érosion utilisant le même élément structurant).

Les premières approches de morphologie mathématique [91] ont concerné des images binaires, c'est-à-dire des images issues d'une classification. Dans ces approches, chaque pixel d'une image numérique binaire est classifié comme appartenant à une des formes à analyser (classe des objets) ou non (classe du fond). Le résultat de cette classification est une image binaire I, i.e. $I_i \in \{0, 1\}$, $\forall i \in \{1, ..., n\}$.

Une opération morphologique peut être vue comme une modification de cette classification en utilisant le voisinage défini par l'élément structurant. Ce voisinage peut également

être vu comme un noyau maxitif binaire $(\nu_j)_{j=1,...,m}$ qui peut être placé sur les pixels $k \in \{1, ..., n\}$ de l'image binaire. On notera alors $(\nu_j^k)_{j=1,...,m}$ ce voisinage placé sur le pixel k. La modification de classification dite de dilatation va résulter en l'image :

$$DI_k = \mathbb{1}_{\{j|\nu_j^k=1\}\cap\{j|I_j=1\}\neq\emptyset},\ \forall k \in \{1, ..., n\}. \tag{III.29}$$

On donne une valeur de 1 à l'image en k s'il existe un pixel de l'élément structurant placé en k qui ait un point d'intersection avec l'image binaire à dilater, et 0 sinon.

La modification de classification dite d'érosion va résulter en l'image :

$$EI_k = \mathbb{1}_{\{j|\nu_j^k=1\}\subseteq\{j|I_j=1\}},\ \forall k \in \{1, ..., n\}. \tag{III.30}$$

On donne une valeur de 1 à l'image en k si tous les pixels de l'élément structurant placé en k font partie de l'image binaire à éroder, et 0 sinon.

L'extension que nous proposons ici est une morphologie floue, car elle considère l'image comme une fonction et l'élément structurant comme un noyau maxitif non binaire. Elle correspond à la définition 1 de [11].

Dans cette approche, l'image I est à niveau de gris, i.e. $\forall i \in \{1, ..., n\}$, $I_i \in \mathbb{R}^+$, et on peut en dégager des images binaires de niveau γ par la donnée des image I^γ définies par :

$$\begin{aligned} I_i^\gamma &= 1, \quad \text{si } I_i \geq \gamma, \\ I_i^\gamma &= 0, \quad \text{sinon.} \end{aligned}$$

On définit également, pour chaque pixel i, un noyau maxitif ς^i qui correspond à l'appartenance d'un élément $\omega \in \Omega$ au voisinage du pixel i. Ses α-coupes sont définies par $\varsigma_\alpha^i = \{\omega \in \Omega | \varsigma^i(\omega) \geq \alpha\}$.

L'élément structurant placé sur un pixel k est modélisé par un noyau maxitif ν^k qui correspond à l'appartenance d'un élément $\omega \in \Omega$ à l'élément structurant placé au pixel k. Ses α-coupes sont définies par $\nu_\alpha^k = \{\omega \in \Omega | \nu^k(\omega) \geq \alpha\}$.

Regardons la généralisation de la dilatation pour l'image I^γ de niveau γ. Pour un élément structurant placé en k, on regarde, pour un des pixels i de I^γ, si l'intersection des α-coupes de l'élément structurant avec les α-coupes du pixel est non vide. On donne la valeur 1 à la dilatation au niveau α de I^γ.

Cela se traduit mathématiquement par :

$$DI_k^{\gamma\alpha} = \sup_{i=1,...,n} \Big[\sup_{\omega\in\Omega} \big[\mathbb{1}_{\varsigma_\alpha^i}(\omega) \mathbb{1}_{\nu_\alpha^k}(\omega) \big] I_i^\gamma \Big],\ \forall k \in \{1, ..., n\}. \tag{III.31}$$

L'intégration des toutes les valeurs $DI_k^{\gamma\alpha}$, suivant les niveaux γ et α, pour tout pixel $k \in \{1, ..., n\}$, nous donne l'extension floue de la morphologie que nous considérons ici :

$$\begin{aligned} DI_k &= \int_0^{+\infty} \int_0^1 DI_k^{\gamma\alpha} d\alpha d\gamma, \\ &= \int_0^{+\infty} \int_0^1 \sup_{i=1,...,n} \Big[\sup_{\omega\in\Omega} \big[\mathbb{1}_{\varsigma_\alpha^i}(\omega) \mathbb{1}_{\nu_\alpha^k}(\omega) \big] I_i^\gamma \Big] d\alpha d\gamma, \\ &= \int_0^{+\infty} \sup_{i=1,...,n} \Big[\sup_{\omega\in\Omega} \big[\int_0^1 \mathbb{1}_{\varsigma_\alpha^i}(\omega) \mathbb{1}_{\nu_\alpha^k}(\omega) d\alpha \big] I_i^\gamma \Big] d\gamma. \end{aligned}$$

Or, $\forall \alpha \in [0, ..., 1]$ et $\forall \omega \in \Omega$, $\mathbb{1}_{\varsigma^i_\alpha}(\omega)\mathbb{1}_{\nu^k_\alpha}(\omega) = \mathbb{1}_{\varsigma^i_\alpha \cap \nu^k_\alpha}(\omega)$. Cette indicatrice est donc égale à 1 jusqu'à un certain α_0, car les coupes de niveaux de distributions de possibilités sont emboitées. α_0 peut être nul, si ω n'est pas du tout dans le support de l'intersection des distributions de possibilité ν^k et ς^i. Ainsi, $\int_0^1 \mathbb{1}_{\varsigma^i_\alpha \cap \nu^k_\alpha}(\omega)d\alpha = \int_0^{\alpha_0} \mathbb{1}_{\varsigma^i_\alpha \cap \nu^k_\alpha}(\omega)d\alpha = \alpha_0 = \sup\{\alpha | \omega \in \varsigma^i_\alpha \cap \nu^k_\alpha\}$. On sait (cf. [30]) qu'une telle expression est identifié à un sous-ensemble flou défini par $min(\varsigma^i(\omega), \nu^k(\omega))$. Pour résumer :

$$\int_0^1 \mathbb{1}_{\varsigma^i_\alpha}(\omega)\mathbb{1}_{\nu^k_\alpha}(\omega)d\alpha = min(\varsigma^i(\omega), \nu^k(\omega)).$$

On obtient donc :

$$DI_k = \int_0^{+\infty} \sup_{i=1,...,n} \Big[\sup_{\omega\in\Omega} min(\varsigma^i(\omega), \nu^k(\omega))I_i^\gamma\Big] d\gamma.$$

Posons

$$\pi_i^k = \sup_{\omega\in\Omega} min(\varsigma^i(\omega), \nu^k(\omega)).$$

π_i^k est une distribution de possibilité. En effet,

$$\pi_k^k = \sup_{\omega\in\Omega} min(\varsigma^k(\omega), \nu^k(\omega)) = \nu^k(\omega_k) = \varsigma^k(\omega_k) = 1.$$

La mesure de possibilité associée s'écrit donc, pour tout $A \subseteq \{1, ..., n\}$

$$\Pi_{\pi^k}(A) = \sup_{i\in A} \pi_i^k = \sup_{i=1,...,n} \mathbb{1}_A(i)\pi_i^k.$$

Au vu de la définition de I^γ, on peut retrouver facilement, de la même manière que dans le calcul II.4, qui tranforme une intégrale de Choquet continue en intégrale de Choquet discrète, que le dilaté en k peut s'écrire exactement comme l'expression (III.28), i.e. l'intégrale de Choquet de l'image I par rapport à la mesure de possibilité Π_{π^k} :

$$DI_k = \sum_{i=1}^{n} (I_{\sigma(i)} - I_{\sigma(i-1)})\Pi_{\pi^k}(A_i), \tag{III.32}$$

avec $A_i := \{\sigma(i), ..., \sigma(n)\}$, $\sigma(0) = 0$ et σ la permutation sur Ω, telle que $I_{\sigma(1)} \leq ... \leq I_{\sigma(n)}$. Les A_i sont des coalitions de pixels de l'image basés sur l'ordre σ, et par convention, $I_0 = 0$.

L'image érodée s'obtient à partir de la mesure de nécessité N_{π^k} associée à π^k, i.e.

$$EI_k = \sum_{i=1}^{n} (I_{\sigma(i)} - I_{\sigma(i-1)})N_{\pi^k}(A_i). \tag{III.33}$$

La démonstration pour l'érosion est similaire à la démonstration pour la dilatation.

III.4.3 Transformation géométrique rigide maxitive

Les méthodes de recalage ou plutôt les transformations rigides, exposées en section I.3.4, présentent un inconvénient majeur : sous l'hypothèse optimiste de mouvements parfaitement rigides, les passages C/D nécessaires à cette méthode induisent des erreurs dont l'influence sur l'image transformée n'est pas mesurée. Les noyaux choisis pour réaliser

ces passages C/D sont généralement choisis pour minimiser ces erreurs mais aussi pour induire un algorithme simple.

Comme nous l'avons vu en section I.3.4, la transformation rigide sommative d'une image $(I_i)_{i=1,...,n}$ en une image $(I'_k)_{k=1,...,n}$ s'écrit, pour tout $k \in \{1, ..., n\}$,

$$I'_k = \sum_{i=1}^{n} I_i t_i^k, \tag{III.34}$$

où

$$t_i^k = \int_\Omega \eta_i^\omega t^{-1} \nu^k(\omega) d\omega, \tag{III.35}$$

avec $(\nu^k)_{k=1,...,n}$, les noyaux d'échantillonnage de l'image transformée I' ; t^{-1}, la transformation inverse de la transformation rigide. Cette transformation inverse est ici non plus appliqué aux points d'échantillonnage de l'image transformée directement, mais aux poids des noyaux d'échantillonnage de l'image transformée $(\nu^k)_{k=1,...,n}$; $(\eta^\omega)_{\omega\in\Omega}$ sont les noyaux de reconstruction de l'image originale I. Nous renvoyons au schéma de la figure I.8 pour une visualisation de cette approche.

Dans cette transformation rigide sommative, les coefficients de la transformation en un $k \in \{1, ..., n\}$, qui sont les $(t_i^k)_{i=1,...,n}$, peuvent être interprétés comme l'espérance mathématique, suivant le noyau sommatif $t^{-1}\nu^k$, de la valeur des poids de reconstruction $(\eta_i^\omega)_{\omega\in\Omega}$ en i. Le calcul des coefficients de la transformation rigide, en expression (III.35), est illustrée sur la figure III.6. Ce calcul est une procédure sommative :

$$t_i^k = \mathbb{E}_{t^{-1}\nu^k}(\eta_i^\bullet), \tag{III.36}$$

où le $\bullet$ représente la variable muette.

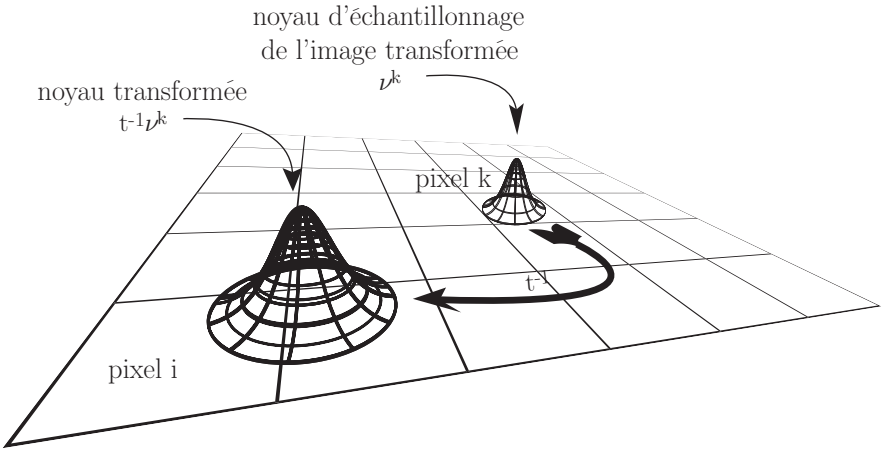


Fig. III.6 : Illustration du calcul des coefficients sommatifs : $(t_i^k)_{i=1,...,n}$

Nous proposons ici une version maxitive de cette transformation. Soient $(\pi^k)_{k=1,...,n}$, la famille de noyaux maxitifs d'échantillonnage de l'image transformée et $(\mu^\omega)_{\omega\in\Omega}$, la famille de noyaux maxitifs de reconstruction de l'image originale. En appliquant la transformation inverse aux noyaux maxitifs d'échantillonnage, on obtient les noyaux maxitifs $(t^{-1}\pi^k)_{k=1,...,n}$ (au lieu des noyaux sommatifs $(t^{-1}\nu^k)_{k=1,...,n}$). On peut ainsi obtenir des coefficients inférieurs et supérieurs de transformation pour chaque pixel $k \in \{1, ..., n\}$, notés

$(\underline{t_i^k})_{i=1,...,n}$ et $(\overline{t_i^k})_{i=1,...,n}$, par extraction maxitive à partir des poids maxitifs de reconstruction en i : les poids $(\mu_i^\omega)_{\omega\in\Omega}$. On remplace l'extraction sommative, pour calculer des coefficients sommatifs (III.36), par une extraction maxitive, pour calculer des coefficients maxitifs :

$$\overline{t_i^k} = \mathbb{C}_{t^{-1}\pi^k}(\mu_i^\bullet), \tag{III.37}$$

$$\underline{t_i^k} = \mathbb{C}^c_{t^{-1}\pi^k}(\mu_i^\bullet). \tag{III.38}$$

$(\overline{t_i^k})_{i=1,...,n}$ est le noyau supérieur de transformation en $k \in \{1, ..., n\}$, au sens où il est moins spécifique que le noyau $(\underline{t_i^k})_{i=1,...,n}$ car $\forall i \in \{1, ..., n\}$, $\underline{t_i^k} \leq \overline{t_i^k}$. $(\underline{t_i^k})_{i=1,...,n}$ est appelé le noyau inférieur de transformation en $k \in \{1, ..., n\}$.

Pour que le noyau supérieur de transformation soit maxitif, des conditions sur les noyaux maxitifs d'échantillonnage et de reconstruction utilisés doivent être rajoutées. Remarquons, en préambule au théorème qui résume ces conditions, que l'on note $(t^{-1}\pi^k)^\alpha$, les α-coupes de $t^{-1}\pi^k$ et $(t^{-1}\pi^k)^1$, le cœur du noyau maxitif $t^{-1}\pi^k$: $(t^{-1}\pi^k)^\alpha = \{\omega \in \Omega | t^{-1}\pi^k \geq \alpha\}$ et $(t^{-1}\pi^k)^1 = \{\omega \in \Omega | t^{-1}\pi^k = 1\}$. Le terme noyau est également utilisé dans la littérature des sous-ensembles flous pour nommer l'α-coupe de niveau 1. On lui préfère ici le terme de cœur pour éviter les confusions avec les noyaux sommatifs ou maxitifs.

Théorème III.1. *À un k donné, s'il existe un $i_0 \in \{1, ..., n\}$ tel que pour au moins un $\omega_0 \in (t^{-1}\pi^k)^1$, on a $\mu_{i_0}^{\omega_0} = 1$, alors $(\overline{t_i^k})_{i=1,...,n}$ est un noyau maxitif.*

Preuve : Notons d'abord [34] que $\overline{t_{i_0}^k}$ s'écrit :

$$\overline{t_{i_0}^k} = \int_0^1 \sup_{\omega\in\Omega}\{\mu_{i_0}^\omega | t^{-1}\pi^k(\omega) \geq \alpha\} d\alpha.$$

On remarque que $\forall \alpha \in [0, 1]$,

$$\omega_0 \in \{\omega | t^{-1}\pi^k(\omega) \geq \alpha\} = (t^{-1}\pi^k)^\alpha,$$

car $\omega_0 \in (t^{-1}\pi^k)^1 \subseteq (t^{-1}\pi^k)^\alpha$.

Ensuite, on remarque que $\forall \alpha \in [0, 1]$, $\forall \omega \in (t^{-1}\pi^k)^\alpha$, $\mu_{i_0}^{\omega_0} \geq \mu_{i_0}^\omega$, car $\mu_{i_0}^{\omega_0} = 1$ et $\mu_{i_0}^\omega$ est une distribution de possibilité, donc $1 \geq \mu_{i_0}^\omega$. On en déduit que, $\forall \alpha \in [0, 1]$,

$$\mu_{i_0}^{\omega_0} = \sup_{\omega\in\Omega}\{\mu_{i_0}^\omega | t^{-1}\pi^k(\omega) \geq \alpha\}.$$

D'où, $\overline{t_{i_0}^k} = \int_0^1 \mu_{i_0}^{\omega_0} d\alpha = \int_0^1 1 d\alpha = 1$. Donc $(\overline{t_i^k})_{i=1,...,n}$ est un noyau maxitif.

□

Il est à noter que pour tout $\omega_0 \in (t^{-1}\pi^k)^1$, il existe un $i_0 \in \{1, ..., n\}$, tel que $\mu_{i_0}^{\omega_0} = 1$. Mais les conditions du théorème III.1 ne sont pas prises dans ce sens là. Cette condition impose que pour chaque pixel k de l'image d'arrivée, il existe un pixel i de l'image de départ, tel que son cœur ait une intersection avec le cœur de la transformée inverse du pixel k. Ici, nous avons confondu pixel et noyau maxitif associé pour simplifier l'explication. Cette condition est tout à fait raisonnable. Par exemple, le cas où les pixels de l'image d'arrivée sont modélisés par un noyau maxitif dont le cœur est de support supérieur ou égal au pas

d'échantillonnage respecte les conditions du théorème III.1, quels que soient les formes imposées aux pixel de l'image de départ. L'illustration de la figure III.7 peut aider à la compréhension de cette condition.

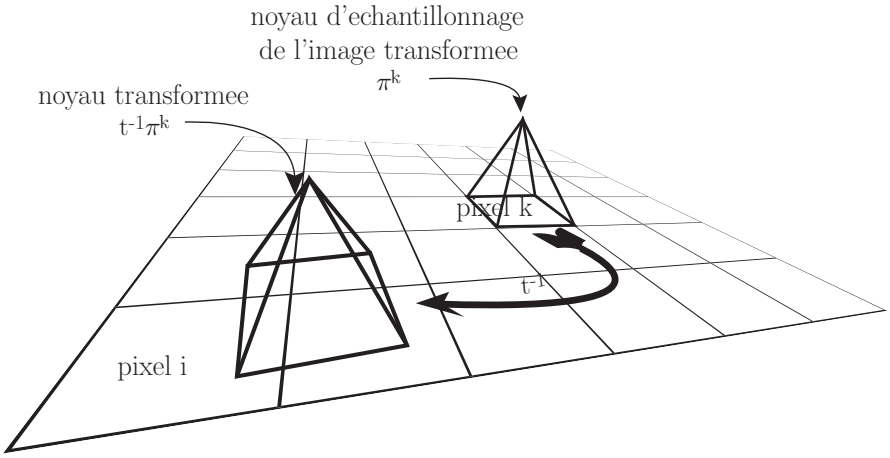


FIG. III.7 : Illustration du calcul des coefficients maxitifs : $(\overline{t_i^k})_{i=1,\dots,n}$

Pour une éventuelle condition de maxitivité d'un noyau inférieur de transformation en k, il faut que, pour un i_0, tous les noyaux maxitifs $\mu_{i_0}^{\omega}$, tels que ω soit hors du support de $t^{-1}\pi^k$, vérifient $\mu_{i_0}^{\omega} = 1$. Cette condition est difficilement réalisable. Il faudrait que pour au moins un pixel i_0, tous les noyaux de reconstruction atteignent leur maximum sur ce point (sauf quelques noyaux maxitifs de reconstruction centrés sur le voisinage $t^{-1}\pi^k$.

Dans notre approche, seul le noyau maxitif supérieur de transformation est utilisé pour réaliser une estimation imprécise de la transformée de l'image I'. Le noyau maxitif inférieur de transformation est plus spécifique que le noyau maxitif supérieur de transformation. En utilisant ce dernier, on réalise donc une estimation plus pessimiste de l'image transformée qu'avec le noyau maxitif de transformation inférieur. On préfère généralement garder le résultat le moins spécifique pour plus de sécurité, donc utiliser $(\overline{t_i^k})_{i=1,\dots,n}$ comme noyau maxitif de transformation.

Ainsi, quand la transformation rigide sommative peut s'écrire :

$$I'_k = \mathbb{E}_{t^k}(I), \tag{III.39}$$

la transformation rigide imprécise maxitive s'écrit :

$$\overline{I'_k} = \mathbb{C}_{\overline{t^k}}(I), \tag{III.40}$$

$$\underline{I'_k} = \mathbb{C}^c_{\overline{t^k}}(I). \tag{III.41}$$

L'estimation imprécise de transformation englobe donc toutes les transformations précises que l'on aurait avec des noyaux sommatifs des familles $(\mathcal{M}(\mu^{\omega}))_{\omega\in\Omega}$ et $(\mathcal{M}(\pi^k))_{k\in\{1,\dots,n\}}$. On ne peut cependant pas garantir qu'après une transformation imprécise et sa transformation inverse imprécise, l'image originale sera dans l'intervalle obtenu.

III.4.4 Détection de contour sur une image par noyau maxitif

III.4.4.1 Principe et définitions

Un contour sur une image est, par définition, une zone de transition franche entre deux zones homogènes. Les points de contours d'une image sont donc des points où les

contrastes sont importants. La méthode la plus utilisée, pour détecter des contours dans une image, consiste à calculer le gradient de cette image et à en déterminer les minima et maxima. Cependant, les contours ne sont pas les seules sources de variation brusque de luminance. Les bruits d'acquisition, de reconstruction ou d'échantillonnage amènent des disparités dans l'image qui peuvent induire des extrema dans la dérivée estimée.

La réduction des effets du bruit et la détection de contour amènent des critères antagonistes pour définir une bonne méthode de détection de contour. L'opérateur de Canny [13, 22] est donné pour optimal pour des critères particuliers, mais d'autres détecteurs de contours existent, qui sont optimaux pour d'autres critères [95]. Même avec de tels filtres optimaux, des discontinuités dues aux bruits persistent que l'on élimine par des méthodes impliquant des seuillages arbitraires. La robustesse de ces méthodes est considérablement réduite par l'arbitraire de ces choix (critères et seuil).

La méthode utilisée, pour retirer de l'image de contour les faux extrema, implique généralement un seuillage par hystérésis qui nécessite le choix de deux valeurs de seuils : s_b et s_h. Les pixels dont la norme du gradient est supérieure à s_b sont, dans un premier temps, présélectionnés (plus le seuil s_b est faible, plus on présélectionne de pixels). Si ce pixel est connecté, par un chemin de pixels de normes supérieures à s_b, à un pixel de norme supérieure à s_h, on sélectionne ce pixel. Tous les autres pixels de l'image de contour sont éliminés, car considérés comme dus au bruit.

La méthode que nous proposons peut être considérée comme basée sur un seuillage naturel implicite qui dépend du niveau de bruit estimé aux points d'échantillonnage. Elle consiste à calculer la dérivée de l'image par l'approche maxitive telle qu'exposée en section III.3.2 et à utiliser cette expression intervalliste de la dérivée pour sélectionner les extrema significatifs.

La dérivée d'une image $I = (I_i)_{i=1,...,n}$, est donnée (cf. section III.3.2) pour tout $i \in \{1, ..., n\}$, par :

$$D\tilde{I}_i = a^-[\mathbb{C}^c_{\pi^{k-}}(I), \mathbb{C}_{\pi^{k-}}(I)] \ominus a^+[\mathbb{C}^c_{\pi^{k+}}(I), \mathbb{C}_{\pi^{k+}}(I)], \tag{III.42}$$

Cette procédure de sélection se déroule comme suit. Supposons que 0 soit inclus dans l'intervalle $D\tilde{I}_i$. Cela signifie qu'il existe une combinaison de noyaux sommatifs d'échantillonnage de reconstruction et d'acquisition, dans les familles de noyaux dominés par les noyaux maxitifs utilisés, qui résulte en une dérivée nulle de l'image au pixel considéré. Les pixels i, tels que $0 \in D\tilde{I}_i$ sont donc écartés. Un point de l'image échantillonnée i est déclaré comme un point de contour s'il maximise la norme moyenne du gradient et si 0 n'appartient pas à l'intervalle d'estimation $D\tilde{I}_i$ de la dérivée de l'image I en i.

III.4.4.2 Expériences

Les expériences que nous présentons ici ont été réalisées en collaboration avec Florence Jacquey et Frédéric Comby [48, 47].

III.4.4.2.1 Expérience sur image réelle

La première expérience vise à comparer qualitativement notre approche à l'approche classique utilisant le filtre de Canny-deriche sur une image réelle, présentée sur la figure III.8. La figure III.9 présente la même image dont le contraste a été rehaussé par l'opération d'accentuation avec gain de 300% réalisée sous photoshop©. Cette image permet de

mieux visualer les contours à extraire de l'image originale. Toutes les expériences et tests sont bien sûr réalisés sur la figure originale (non contrastée).

L'image présentée en figure III.8 est particulièrement intéressante pour étudier des algorithmes de détection de contour. Elle possède des zones où des contours existent mais dont les contrastes ne sont pas très importants comme, par exemple, la grille de la partie haute (zone (a) de la figure III.9), mettant en avant la capacité de détection de l'algorithme. Elle possède des zones où il y a beaucoup de variations et de contrastes dans l'image comme, par exemple, les écailles des poissons (zone (b) de la figure III.9), mettant en avant le comportement du détecteur face au bruit. Elle possède des zones où les contours sont bien visibles bien contrastés comme, par exemple, la queue des poissons (zone (c) de la figure III.9) permettant de vérifier que le détecteur détecte bien des contours simples. Elle possède des zones homogènes comme, par exemple, la glissière (zone (d) de la figure III.9), montrant la sensibilité du détecteur au seuillage.

Fig. III.8 : Poissons (la vraie image)

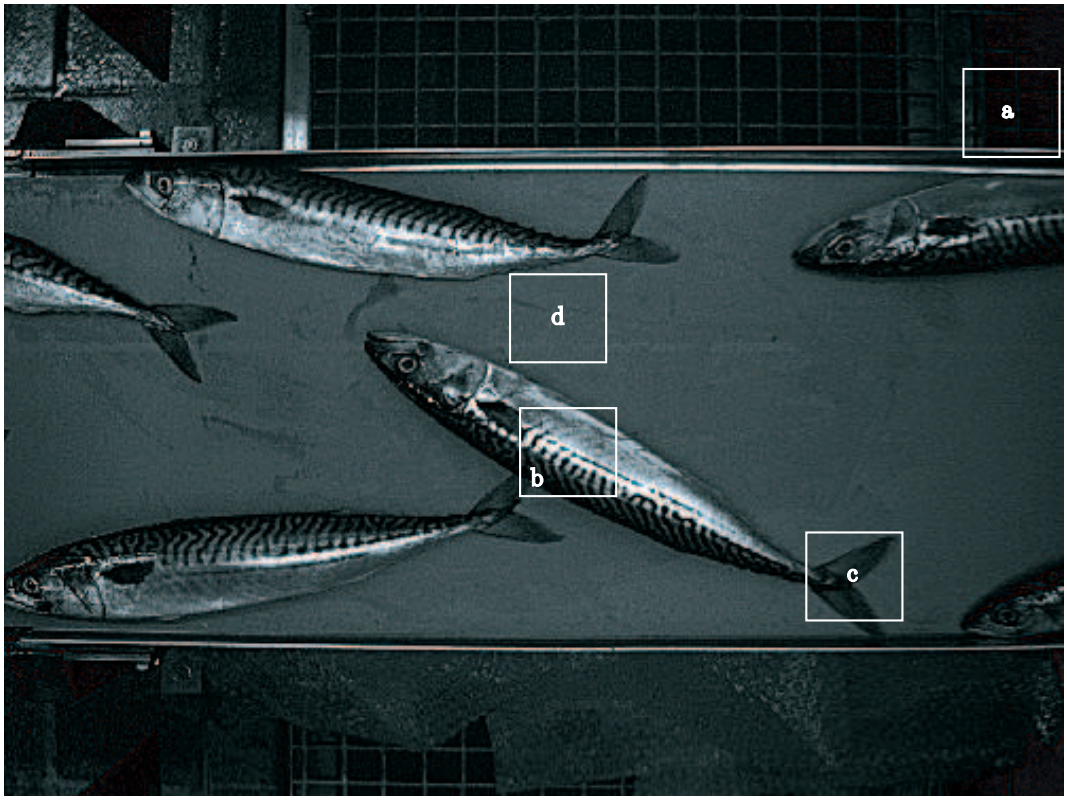


Fig. III.9 : Poissons (l'image contrastée)

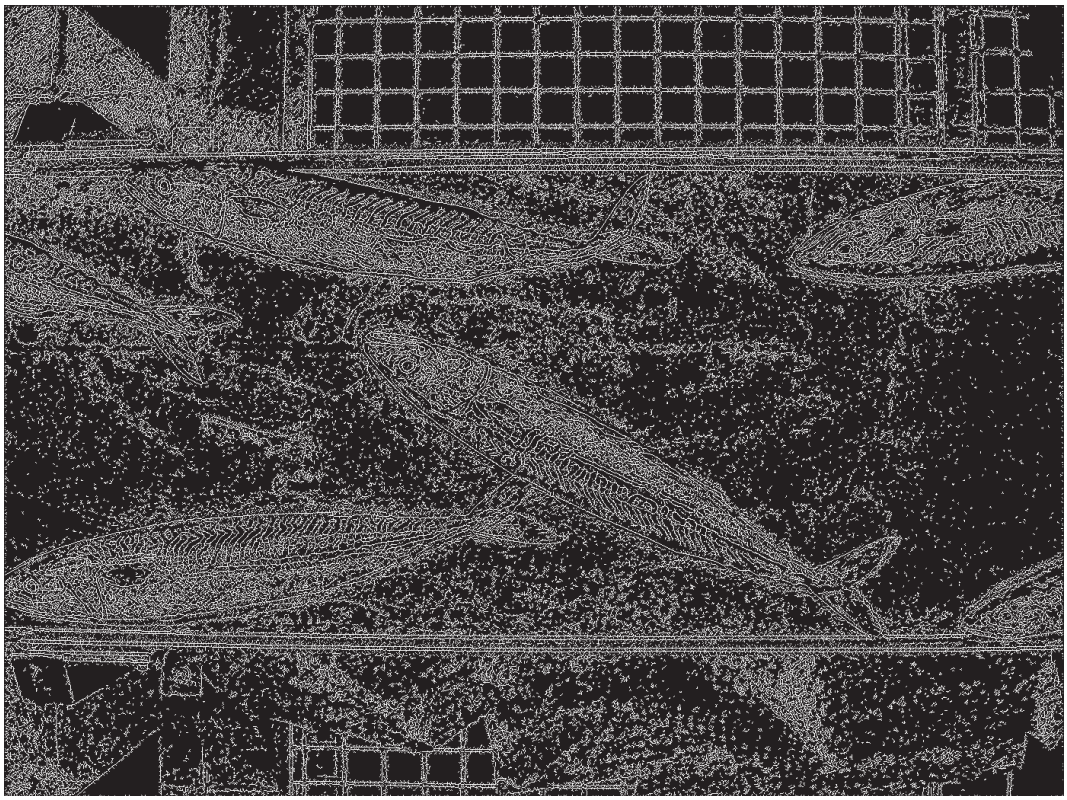


Fig. III.10 : Filtre de Canny-Deriche pour $s_b = 4$ et $s_h = 6$

Fig. III.11 : Filtre de Canny-Deriche pour $s_b = 28$ et $s_h = 30$

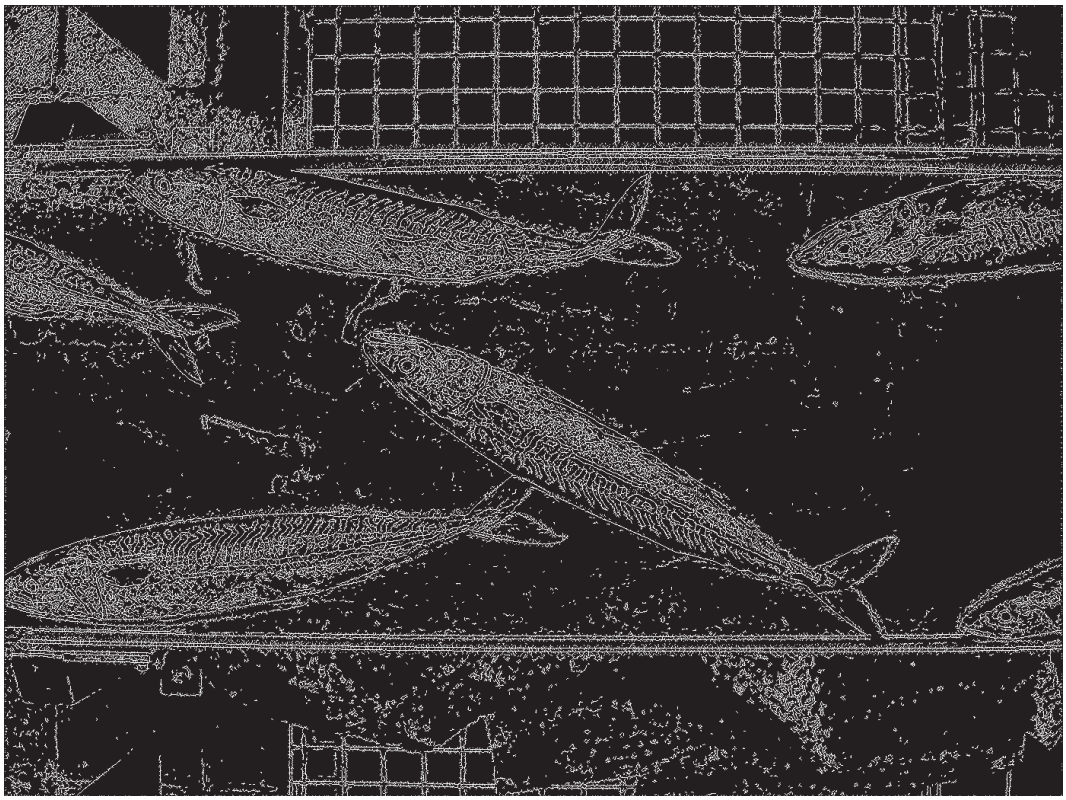


Fig. III.12 : Filtre de Canny-Deriche pour $s_b = 13$ et $s_h = 16$

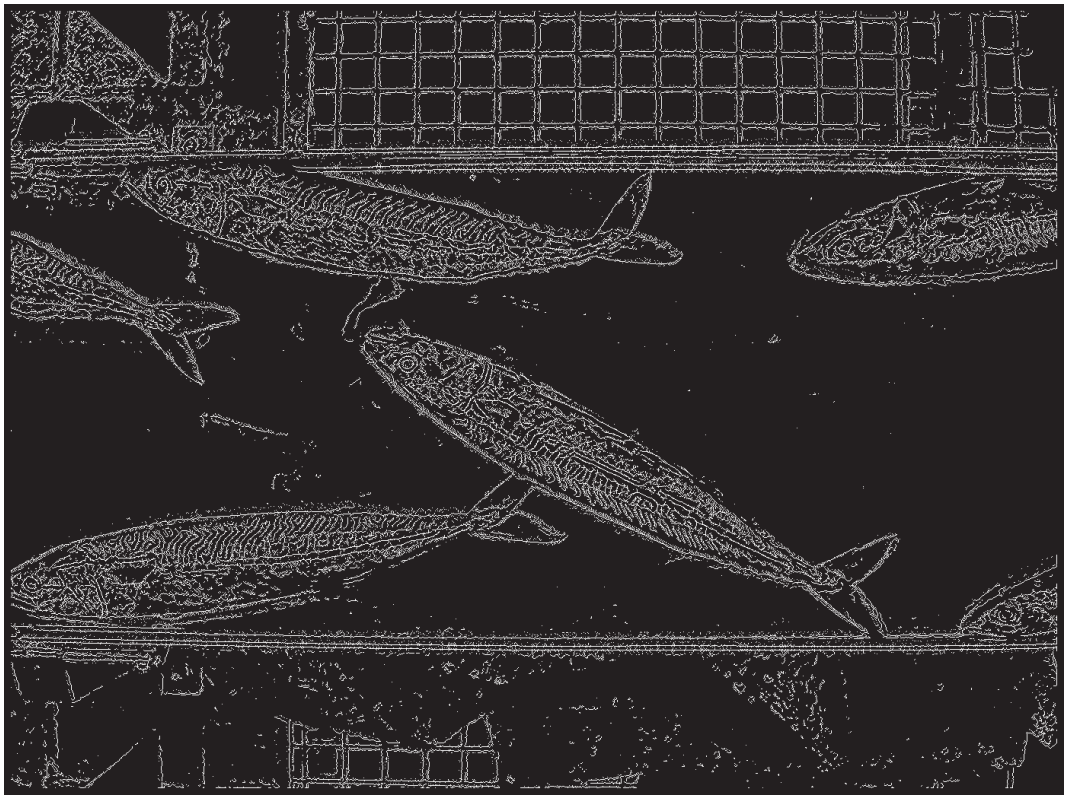


Fig. III.13 : Filtre maxitif

Nous présentons les résultats de ces algorithmes sur toute l'image aux figures III.10, III.11, III.12 et III.13. La figure III.10 représente les contours obtenus avec un filtre de Canny-Deriche, pour des seuils bas et haut choisis à $s_b = 4$ et $s_h = 6$. La figure III.11 représente les contours obtenus avec un filtre de Canny-Deriche, pour des seuils bas et haut choisis à $s_b = 28$ et $s_h = 30$. La figure III.12 représente les contours obtenus avec un filtre de Canny-Deriche, pour des seuils bas et haut choisis à $s_b = 13$ et $s_h = 16$. La figure III.13 représente les contours obtenus avec un filtre maxitif.

Si on observe la détection de contours obtenue avec un couple de seuils bas (figure III.10), les vrais contours sont bien détectés, mais sont aussi détectés les faux contours induits par le bruit. À l'inverse, si on observe la détection de contours obtenue avec un couple de seuils hauts (figure III.11), les faux contours induits par le bruit ne sont pas détectés, mais beaucoup de vrais contours ne le sont pas non plus.

Pour le détecteur de la figure III.12, nous avons empiriquement déterminer les seuils donnant un bon compromis entre détection de contours minimisation des effets du bruit ($s_b = 13$ et $s_h = 16$). Nous présentons sur la figure III.14, quatre images de détails (a1), (b1), (c1) et (d1) résultant en des détections de contours (a2), (b2),(c2) et (d2) obtenues avec le filtre de Canny-Deriche ($s_b = 13$ et $s_h = 16$) et des détections de contours (a3), (b3),(c3) et (d3) obtenues avec l'approche maxitive.

La zone (a1) représente une grille régulière peu contrastée. Les contours de cette grille semblent correctement détectés par notre approche (a3) alors qu'avec l'approche de Canny Deriche (a2), il manque des contours. La zone (b1) représente le corps d'un poisson. Les écailles en sont bien détectées par les deux approches, mais notre approche (b3) est beaucoup moins sensible au bruit (partie ventrale du poisson) que l'approche de Canny-Deriche (b2). La zone (c1) représente la queue de ce même poisson. Notre approche (c3)

semble dessiner des contours plus nette que l'approche de Canny-Deriche (c2). La zone (d1) est une zone homogène de la glissière à poisson. Notre approche (d3) est beaucoup moins sensible au bruit que l'approche de Canny-Deriche (d2). Les seuls points de contours détectés par notre approche correspondent bien à des parties plus foncées de l'image.

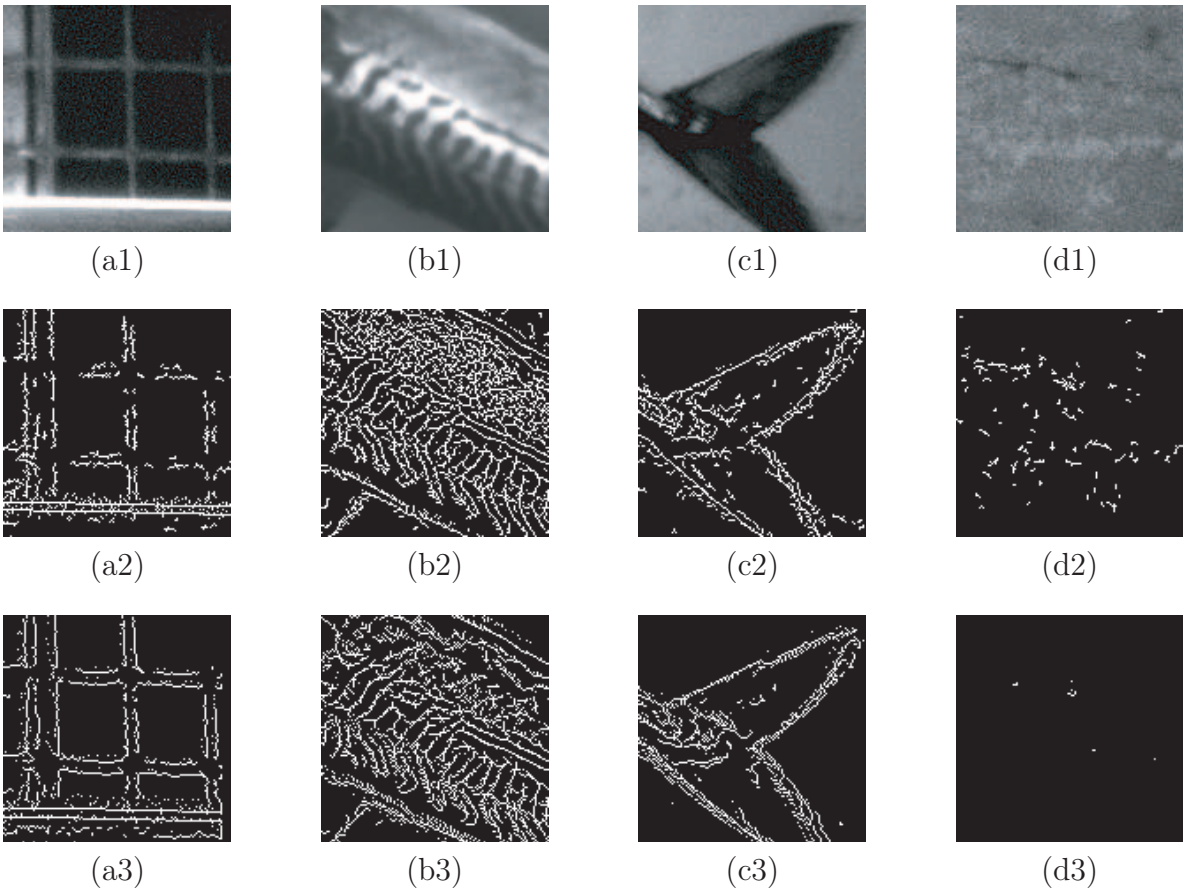
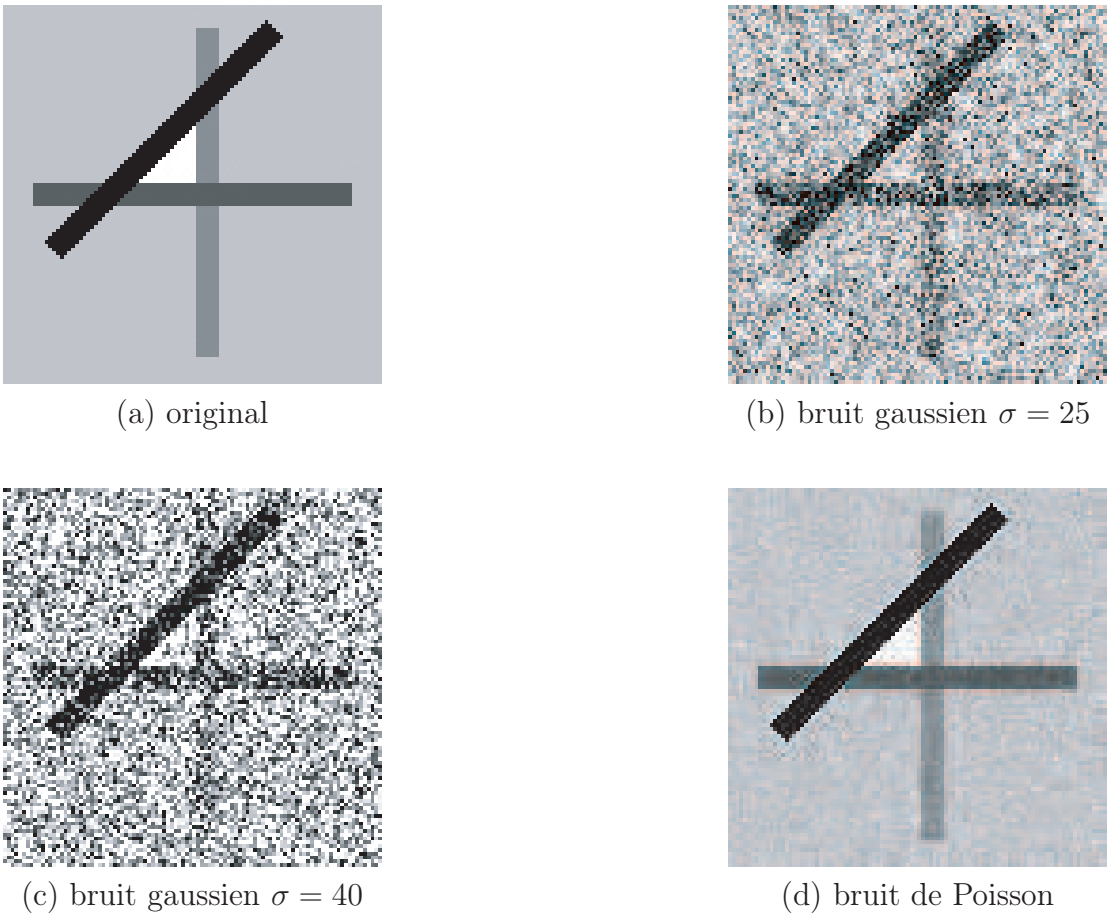


Fig. III.14 : Résultats détaillés des détecteurs de contours

Notre approche semble donc détecter les contours de façon plus robuste que l'approche de Canny-Deriche au sens où elle semble beaucoup moins sensible au bruit. Son utilisation dispense d'un choix arbitraire de seuil.

III.4.4.2.2 Expérience sur image artificielle avec bruit simulé

La seconde expérience vise à comparer, de façon plus objective, notre approche avec celles de Canny-Deriche [22], Shen-Castan [95] et Prewitt, par des expériences sur une image artificielle, présentée sur la figure III.15(a). Cette image a été parasitée par deux types de bruit : un bruit gaussien de variance σ variable (la figure III.15(b) présente l'image parasitée pour une variance égale à 25 et la figure III.15(c) présente l'image parasitée pour une variance égale à 40) et un bruit de Poisson, dont le niveau dépend de l'intensité de l'image (cf. figure III.15(d)). Utiliser un bruit gaussien permet de tester la robustesse de notre approche aux niveaux de bruit. Utiliser un bruit de Poisson permet de tester la robustesse de notre approche à des variations, au sein d'une même image, du niveau de bruit. Les positions des contours, connues car c'est une image artificielle, permettent d'employer des critères objectifs de comparaison.



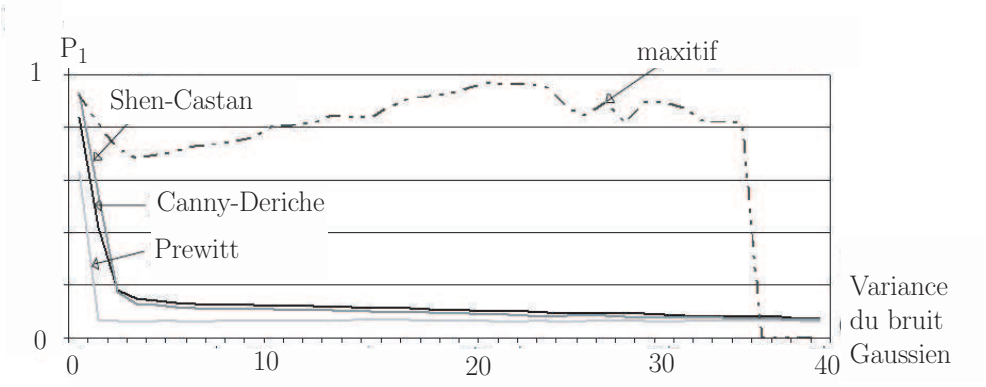
(a) original (b) bruit gaussien $\sigma = 25$

(c) bruit gaussien $\sigma = 40$ (d) bruit de Poisson

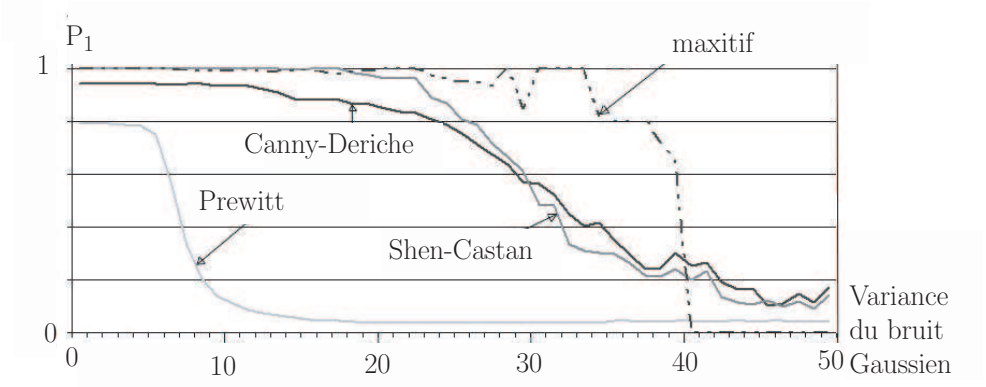
Fig. III.15 : Image artificielle avec bruit gaussien et de Poisson

Les performances des différents détecteurs que nous étudions dans cette expérience sont quantifiées par le paramètre P_1 défini par Fram et Deutsch [23]. P_1 mesure la sensibilité du détecteur en présence de bruit. Il est normalisé par $P_1 = 1$ quand la détection est optimale.

Nous étudions dans un premier temps la sensibilité des différents algorithmes au bruit gaussien. La figure III.16(a) présente le paramètre P_1 obtenu avec les approches de Canny-Deriche, Shen-Castan, Prewitt et l'approche maxitive, en fonction de la variance du bruit gaussien, lorsqu'aucun seuillage n'est effectué. Pour toutes ces approches, exceptée la nôtre, le critère P_1 chute dès que $\sigma \geq 3$. La figure III.16(b) présente le paramètre P_1 obtenu pour ces mêmes approches, lorsqu'un seuillage est effectué. Pour comparer tous les algorithmes de façon optimale, nous avons choisi, pour chaque niveau de bruit et pour chaque filtre, les seuils maximisant le critère P_1. Dans chacune des méthodes, les comportements de toutes les méthodes face au bruit sont améliorés, comme nous le montre la figure III.16(b). On constate que le filtre de Prewitt est le plus sensible au bruit, tandis que les approches de Canny-Deriche et de Shen-Castan ont des comportements similaires. Pour $\sigma < 25$, les contours sont détectés correctement ; pour $25 < \sigma < 40$, le critères P_1 diminue jusqu'à une confusion entre bruits et contours. Avec notre approche, la décroissance, qui commence plutôt à $\sigma = 33$ est douce jusqu'à $\sigma = 40$, puis très nette à 40. À partir d'une variance $\sigma = 40$ (cf. III.15(c)), plus aucun détecteur ne fonctionne.



(a) sans seuillage



(b) avec seuillage

Fig. III.16 : Comparaison des valeurs de P_1 pour une image avec bruit gaussien

Approche	Prewitt	Deriche	Shen	Maxitive
P_1	0.089676	0.14975	0.141	0.69881

Tab. III.2 : Valeurs de P_1 pour des images avec bruit de Poisson

Dans la seconde expérience, nous étudions le comportement des quatre méthodes de détection lorsque l'image est parasitée par un bruit de Poisson III.15(c), c'est-à-dire un bruit non-stationnaire. Les seuils choisis, pour chaque méthode, sont ceux maximisant le critère P_1. Les résultats de cette expérience sont reportés dans le tableau III.2. On constate que l'approche maxitive est 8 fois moins sensible au bruit que l'approche de Prewitt et 5 fois moins que les approches de Canny-Deriche et Shen-Castan.

À cause de la non-stationnarité du bruit, il est difficile de proposer un seuillage, pour chacune des approches de Canny-Deriche, Shen-Castan ou Prewitt, permettant d'éliminer les contours dus au bruit sur toute l'image. Cette remarque explique pourquoi notre méthode (qui finalement adapte le seuillage en chaque point) a de bien meilleures performances que les méthodes basées sur un seuillage fixe.

L'approche maxitive apparait donc très intéressante quand la détection de contours est appliquée sur des images médicales dont le bruit suit généralement un processus de Poisson.

III.5 Statistique maxitive

III.5.1 Estimateur robuste de fonction de répartition par noyau maxitif

III.5.1.1 Principe et définition

Comme nous l'avons noté en section I.4.4, les estimateurs par histogramme (binaire ou flou) de fonction de répartition sont des cas particuliers de l'estimateur à noyau de Parzen Rosenblatt de fonction de répartition. L'estimateur à noyau sommatif est défini, pour tout ω de Ω, par :

$$\hat{F}_\kappa(\omega) = \int_{\{u\in\Omega | u\leq\omega\}} \frac{1}{n}\sum_{i=1}^{n} \kappa(u - x_i)du, \tag{III.43}$$

où κ est un noyau sommatif.

Nous avons montré que cet estimateur de fonction de répartition peut s'écrire sous une forme d'extraction sommative d'informations (I.91). En effet, pour tout ω de Ω, on a :

$$\hat{F}_\kappa(\omega) = \mathbb{E}_{\kappa^\omega}(E_n), \tag{III.44}$$

avec $E_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} H^{x_i}(\omega)$, la fonction de répartition empirique.

La généralisation robuste de cet estimateur que nous proposons s'appuie sur l'extraction maxitive d'informations. Elle consiste à remplacer le noyau sommatif de Parzen Rosenblatt κ par un noyau maxitif π qui représente une famille $\mathcal{M}(\pi)$ de noyaux sommatifs de Parzen Rosenblatt. L'estimateur robuste par noyau maxitif est défini, pour tout ω de Ω, par :

$$\underline{\overline{F}}_\pi(\omega) = [\mathbb{C}^c_{\pi^\omega}(E_n), \mathbb{C}_{\pi^\omega}(E_n)], \tag{III.45}$$

où

$$\mathbb{C}^c_{\pi^\omega}(E_n) = \int_0^\infty N_{\pi^\omega}(\{\omega\in\Omega | E_n(\omega)\geq\alpha\})d\alpha, \tag{III.46}$$

$$\mathbb{C}_{\pi^\omega}(E_n) = \int_0^\infty \Pi_{\pi^\omega}(\{\omega\in\Omega | E_n(\omega)\geq\alpha\})d\alpha. \tag{III.47}$$

D'après le théorème fondamental de l'extraction maxitive d'informations (théorème II.18), tous les estimateurs de Parzen Rosenblatt de fonction de répartition obtenus avec des noyaux sommatifs de $\mathcal{M}(\pi)$ sont dans l'intervalle résultat d'estimation maxitive de fonction de répartition $\underline{\overline{F}}_\pi$. C'est de cette propriété que provient la cohérence théorique de notre approche [60].

On obtient un intervalle d'estimateurs qui prend en compte la méconnaissance du praticien du meilleur noyau à utiliser pour estimer une fonction de répartition.

III.5.1.2 Reformulation pratique de l'estimateur maxitif de fonction de répartition

Regardons pratiquement le calcul de cet estimateur robuste de fonction de répartition. Pour cela, transformons quelque peu E_n, la fonction de répartition empirique. E_n peut

s'exprimer, pour tout ω dans Ω, comme :

$$E_n(\omega) = \sum_{i=1}^{n} \frac{i}{n} \mathbb{1}_{[x_{(i)}, x_{(i+1)}]}, \tag{III.48}$$

où (.) est une permutation sur l'ordre des observations telle que $x_{(1)} \leq ... \leq x_{(n)}$. Ainsi, la borne supérieure de l'estimateur maxitif de fonction de répartition (III.47), qui est l'intégrale de Choquet de E_n par rapport à la mesure de possibilité Π_{π^ω}, peut, grâce à une adaptation simple du calcul II.4, être réécrite comme :

$$\mathbb{C}_{\pi^\omega}(E_n) = \frac{1}{n} \sum_{i=1}^{n} \Pi_{\pi^\omega}(\{u \in \Omega : E_n(u) \geq \frac{i}{n}\}).$$

D'après la réécriture (III.48) de E_n, on peut facilement voir que $\{u \in \Omega : E_n(u) \geq \frac{i}{n}\} = \{u \in \Omega : u \geq x_{(i)}\}$ et donc,

$$\mathbb{C}_{\pi^\omega}(E_n) = \frac{1}{n} \sum_{i=1}^{n} \Pi_{\pi^\omega}(\{u \in \Omega : u \geq x_{(i)}\}).$$

Une somme ne dépendant pas de l'ordre des indices des éléments à sommer, on en déduit que :

$$\mathbb{C}_{\pi^\omega}(E_n) = \frac{1}{n} \sum_{i=1}^{n} \Pi_{\pi^\omega}(\{u \in \Omega : u \geq x_i\}).$$

Par des développements similaires sur la borne inférieure de l'estimateur maxitif de fonction (III.46), on obtient :

$$\mathbb{C}^c_{\pi^\omega}(E_n) = \frac{1}{n} \sum_{i=1}^{n} \Big(1 - \Pi_{\pi^\omega}(\{u \in \Omega : u < x_i\})\Big),$$

$$\mathbb{C}_{\pi^\omega}(E_n) = \frac{1}{n} \sum_{i=1}^{n} \Big(1 - N_{\pi^\omega}(\{u \in \Omega : u < x_i\})\Big).$$

Dans [30, 34], il est montré que les fonctions, définies pour tout $v \in \Omega$, par :

$$\underline{F_{\pi^\omega}}(v) = N_{\pi^\omega}(\{u \in \Omega : u < v\}), \text{ et}$$
$$\overline{F_{\pi^\omega}}(v) = \Pi_{\pi^\omega}(\{u \in \Omega : u < v\}),$$

sont les fonctions de répartitions inférieures et supérieures de l'ensemble des fonctions de répartitions associées aux noyaux sommatifs de $\mathcal{M}(\pi^\omega)$. Ces fonctions de répartition inférieure et supérieure forment une p-box [37] et vérifient (cf. [34]) :

$$\underline{F_{\pi^\omega}}(u) = \begin{cases} 0 & \text{si } u < \omega, \\ 1 - \pi^\omega(u) & \text{sinon,} \end{cases} \text{ et } \overline{F_{\pi^\omega}}(u) = \begin{cases} \pi^\omega(u) & \text{si } u < \omega, \\ 1 & \text{sinon.} \end{cases}$$

On obtient donc les bornes suivantes à notre estimateur :

$$\mathbb{C}^c_{\pi^\omega}(E_n) = \frac{1}{n} \sum_{i=1}^{n} \Big((1 - \pi^\omega(x_i)) \mathbb{1}_{[\omega \geq x_i]}\Big), \tag{III.49}$$

$$\mathbb{C}_{\pi^\omega}(E_n) = \frac{1}{n} \sum_{i=1}^{n} \Big(\pi^\omega(x_i) \mathbb{1}_{[\omega \leq x_i]} + \mathbb{1}_{[x > x_i]}\Big). \tag{III.50}$$

Cette reformulation est particulièrement intéressante car elle aboutit à un calcul, pour chacune des bornes $\mathbb{C}^c_{\pi^\omega}(E_n)$ et $\mathbb{C}_{\pi^\omega}(E_n)$, dont la complexité est la même que celle de $\hat{F}_\kappa$.

III.5.1.3 Granularité du noyau maxitif d'estimation et imprécision de l'estimation

La granularité, définie en section II.5, est un indice d'imprécision du noyau maxitif ainsi que de l'extraction maxitive d'informations. Au travers du cas particulier d'extraction maxitive d'informations, qu'est l'estimateur robuste de fonction de répartition (III.45), nous pouvons justifier cette affirmation.

En effet, l'écart entre la borne supérieure et la borne inférieure de notre estimateur robuste en $\omega \in \Omega$, noté

$$\varepsilon_\pi(\omega) = \mathbb{C}_{\pi^\omega}(E_n) - \mathbb{C}^c_{\pi^\omega}(E_n), \tag{III.51}$$

marque l'imprécision de l'estimateur en ω. La moyenne de cet écart sur Ω est égale à la granularité du noyau maxitif π utilisé.

Théorème III.2. *Pour tout noyau maxitif π,*

$$\Gamma(\pi) = \int_\Omega \varepsilon_\pi(\omega) d\omega, \tag{III.52}$$

où Γ est la granularité définie par (II.38) et ε_π est la fonction d'écart entre les bornes de l'estimateur robuste de fonction de répartition par noyau maxitif définie par (III.51).

Preuve : Partons de la partie droite de l'égalité (III.52) à démontrer. D'après les expressions (III.49) et (III.50) des bornes inférieures et supérieures, on a

$$\varepsilon_\pi(\omega) = \frac{1}{n} \sum_{i=1}^{n} \big(\pi^\omega(x_i) - \mathbb{1}_{\omega=x_i}\big),$$

d'où,

$$\int_\Omega \varepsilon_\pi(\omega) d\omega = \frac{1}{n} \sum_{i=1}^{n} \int_\Omega \big(\pi^\omega(x_i) - \mathbb{1}_{\omega=x_i}\big) d\omega,$$

Or $\forall i \in \{1, ..., n\}$, $\int_\Omega \mathbb{1}_{\omega=x_i} d\omega = 0$, donc

$$\begin{aligned}
\int_\Omega \varepsilon_\pi(\omega) d\omega &= \frac{1}{n} \sum_{i=1}^{n} \int_\Omega \pi^\omega(x_i) d\omega, \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_\Omega \pi(\omega - x_i) d\omega, \\
&= \frac{1}{n} \sum_{i=1}^{n} \Gamma(\pi), \\
&= \Gamma(\pi).
\end{aligned}$$

□

Ce théorème est donc très intéressant pour justifier du choix de l'indice de granularité comme indice d'imprécision en extraction maxitive d'informations. En effet, il est ici un marqueur de l'imprécision moyenne de l'estimateur robuste (III.45).

III.5.1.4 Expérience

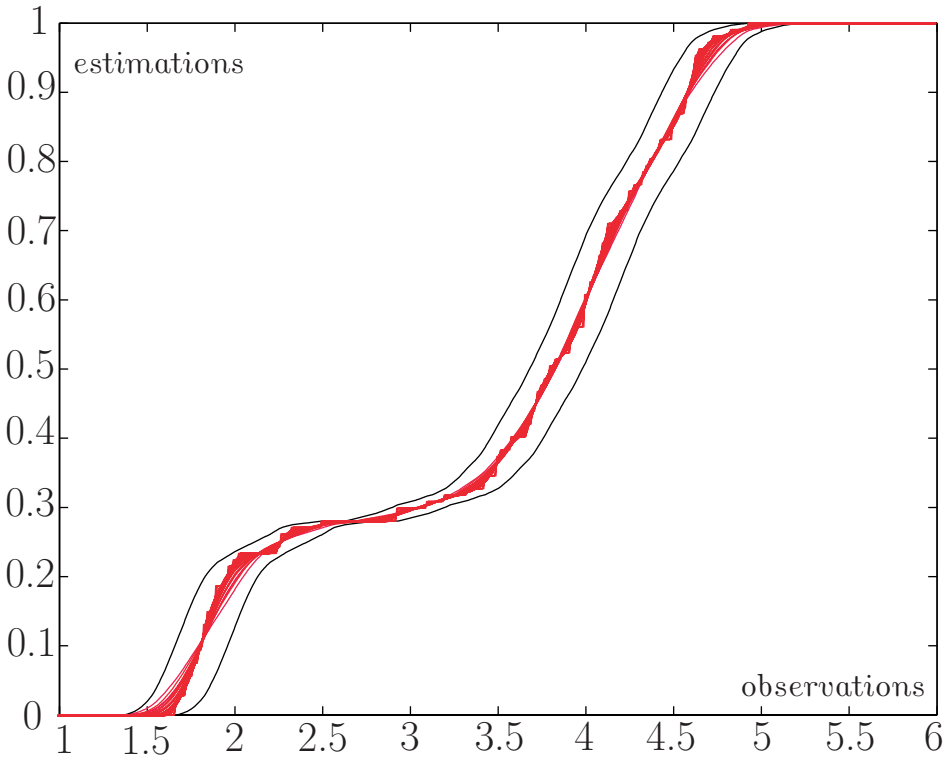


Fig. III.17 : Estimateur robuste de fonction de répartition

Illustrons la cohérence de notre approche par rapport aux estimateurs à noyau de Parzen Rosenblatt par l'expérience suivante. À partir de 107 observations de la durée (en minutes) des éruptions du Old Faithful geyser, situé dans le Yellowstone National Park[2], nous effectuons une estimation maxitive (III.45), via le noyau maxitif triangulaire T, de largeur de bande $\Delta = 0.3$ (cf. expression II.17). Parallèlement, on effectue des estimations précises de la fonction de répartition sous-jacente à ces observations par la méthode de Parzen Rosenblatt avec les noyaux sommatifs uniforme, Epanechnikov, triweight et cosinus pour différentes largeurs de bande inférieures ou égales à Δ. Le noyau maxitif triangulaire utilisé domine donc tous les noyaux sommatifs considérés. L'estimation imprécise (en noire sur la figure III.17) obtenue avec le noyau maxitif triangulaire encadre bien les estimations précises (en rouge sur la figure III.17) obtenues avec les noyaux sommatifs uniforme, Epanechnikov, triweight et cosinus. La figure III.17 est une illustration du fait que tout estimateur à noyau sommatif (de Parzen Rosenblatt) est inclus dans un estimateur robuste à noyau maxitif qui domine son noyau sommatif. Cette propriété est mathématiquement prouvée par le théorème II.18.

[2]Cet exemple, tiré de [96], est un ensemble de données très populaire en statistiques non paramétriques.

III.5.2 Estimateur robuste de densité de probabilité

III.5.2.1 Principe et définition

Nous avons déjà remarqué, en section I.4, que les estimateurs par histogramme (binaire ou flou) de densité de probabilité sont des cas particuliers de l'estimateur à noyau de Parzen Rosenblatt de densité de probabilité. L'estimateur de densité de probabilité à noyau de Parzen Rosenblatt est défini, pour tout ω de Ω, par :

$$\hat{f}_\kappa(\omega) = \frac{1}{n}\sum_{i=1}^{n}\kappa(\omega - x_i), \tag{III.53}$$

où κ est un noyau sommatif.

Dans la section précédente, nous avons montré comment l'outil équivalent à cet estimateur, pour les fonctions de répartition, pouvait être étendue à l'approche maxitive et le rendre plus robuste à des défauts de connaissance du noyau sommatif à utiliser.

Dans cette section, l'extension de l'estimateur (III.53) aux noyaux maxitifs n'est pas aussi direct que l'extension (III.45). En effet, la réécriture de cet estimateur sous forme d'extraction sommative d'informations, donnée par

$$\hat{f}_\kappa(\omega) = \mathbb{E}_{\kappa^\omega}(e_n), \tag{III.54}$$

où e_n est la distribution empirique, ne permet pas un passage direct aux opérations de l'extraction maxitive d'informations. L'intégrale de Choquet n'est définie que pour des fonctions bornées, or la distribution empirique, définie par $e_n = \sum_{i=1}^{n}\delta^{x_i}$, n'est pas une fonction bornée. C'est une distribution au sens de Schwartz [88].

Nous proposons de contourner cette difficulté en passant par la reformulation de l'estimateur de densité de probabilité $\hat{f}_\kappa$ de l'expression (I.101) :

$$\hat{f}_\kappa(\omega) = a^-\mathbb{E}_{d\kappa^{\omega-}}(E_n) - a^+\mathbb{E}_{d\kappa^{\omega+}}(E_n), \tag{III.55}$$

où les éléments a^-, a^+, $d\kappa^{\omega-}$ et $d\kappa^{\omega+}$, sont définis de telle sorte que $d\kappa^\omega = a^+d\kappa^{\omega+} - a^-d\kappa^{\omega-}$ où $d\kappa^\omega$ est le noyau dérivé de κ^ω. Ce type de décomposition est détaillée au théorème I.9 dans la partie filtrage sommatif en section I.3.5.

Le théorème suivant nous permet de définir des bornes d'estimation de la densité de probabilité.

Théorème III.3. *Soit κ un noyau sommatif. Soient π^+ et π^- deux noyaux maxitifs tels que $d\kappa^- \in \mathcal{M}(\pi^-)$ et $d\kappa^+ \in \mathcal{M}(\pi^+)$. Pour tout $\omega \in \Omega$,*

$$\hat{f}_\kappa(\omega) \in a^+\underline{F}_{\pi^+}(\omega) \ominus a^-\underline{F}_{\pi^-}(\omega), \tag{III.56}$$

où $\ominus$ est l'extension de la soustraction aux intervalles.

Preuve : D'aprés (III.55) et (III.44), on a :

$$\widehat{f}_\kappa(x) = a^-\hat{F}_{d\kappa^-}(\omega) - a^+\hat{F}_{d\kappa^+}(\omega).$$

En s'appuyant sur les résultats de [60] présentés dans la section III.5.1, et le théorème fondamental de l'extraction maxitive d'informations, on a :

$$\hat{F}_{d\kappa^-}(\omega) \in \underline{F}_{\pi^-}(\omega), \text{ et}$$
$$\hat{F}_{d\kappa^+}(\omega) \in \overline{F}_{\pi^+}(\omega).$$

D'où ,

$$\hat{f}_\kappa(\omega) \in a^+\overline{F}_{\pi^+}(\omega) \ominus a^-\overline{F}_{\pi^-}(\omega),$$

□

Les bornes inférieures et supérieures des intervalles $\overline{F}_{\pi^+}(\omega)$ et $\overline{F}_{\pi^-}(\omega)$ peuvent être calculées simplement grâce aux formules (III.49) et (III.50). Ce travail a été réalisé en collaboration avec Bilal Nehme [69].

III.5.2.2 Expérimentations

Nous proposons, dans cette section, deux expérimentations illustrant tant les qualités que les défauts de la méthode d'estimation que nous proposons. 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 variance 4 centrée en 8 par une loi normale de variance unitaire et centrée en 3. Nous avons choisi d'utiliser, comme noyau sommatif de référence, le noyau d'Epanechnikov car c'est le plus utilisé en estimation de Parzen Rosenblatt d'une part, et c'est aussi celui qui a le comportement le plus singulier dans l'approche que nous proposons.

Dans la première expérience, nous avons simulé 1000 observations $(x_1, ..., x_{1000})$ de cette loi bimodale simmulée. Nous avons réalisé l'estimation de Parzen-Rosenblatt précise en utilisant un noyau d'Epanechnikov, avec une largeur de bande $\Delta = 0.8$. Ce noyau sommatif est obtenu, via la relation (I.22), à partir de l'expression du noyau d'Epanechnikov qui se trouve dans le tableau II.1. Nous avons, d'autre part, réalisé l'estimation imprécise de cette distribution en dominant le noyau dérivé, comme indiqué par les hypothèses du théorème III.3 grâce à la transformation objective (cf. section II.3.4.3.1) d'une part (Figure III.18) et à la transformation subjective (cf. section II.3.4.3.2) d'autre part (Figure III.19). Nous avons reporté sur chaque figure la courbe de la densité simulée.

On peut voir, sur la Figure III.18, que l'utilisation de la domination la plus spécifique conduit à une estimation inférieure égale à l'estimation par noyau sommatif. Cette propriété est une des particularités du noyau d'Epanechnikov. Par contre, l'examen de la Figure III.19 montre qu'une domination moins spécifique des noyaux de dérivation conduit à une estimation précise qui semble à mi-chemin des bornes de l'estimation imprécise.

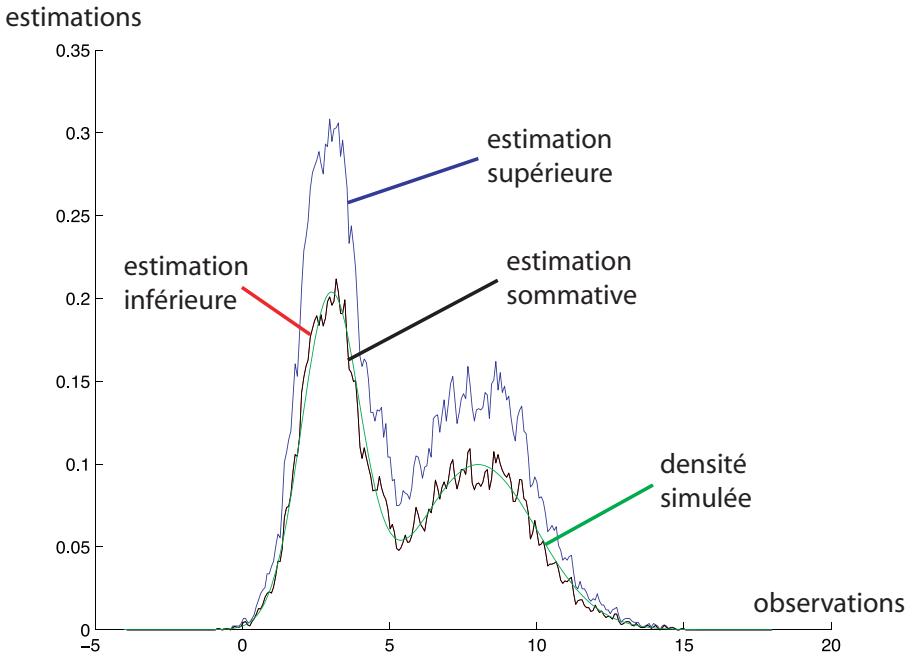


Fig. III.18 : estimation précise et imprécise, domination avec la transformation objective

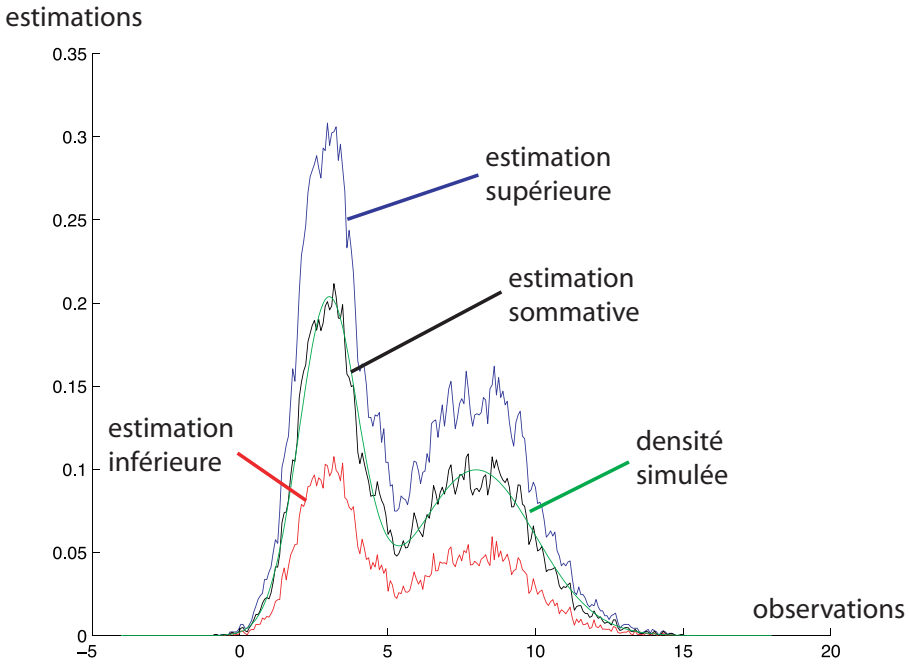


Fig. III.19 : estimation précise et imprécise, domination avec la transformation subjective

La seconde expérience tend à mettre en évidence la robustesse induite par l'utilisation d'une estimation imprécise. Nous proposons de caractériser cette robustesse par deux types d'indice. Le premier montre la corrélation entre la taille de l'intervalle $a^{+}\overline{\underline{F}}_{\pi^{+}}(\omega) \ominus a^{-}\overline{\underline{F}}_{\pi^{-}}(\omega)$ d'une part et la variance d'estimation de $\hat{f}_{\kappa}(\omega)$ d'autre part. Le second montre

l'aptitude de l'estimation imprécise $a^+\overline{\underline{F}}_{\pi^+}(\omega) \ominus a^-\overline{\underline{F}}_{\pi^-}(\omega)$ obtenue avec une seule expérience à contenir la vraie densité $f(\omega)$.

L'expérience se déroule de la manière suivante. On effectue 500000 tirages indépendants 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 estimation précise $\hat{f}_\kappa(\omega)$ et d'une estimation imprécise $a^+\overline{\underline{F}}_{\pi^+}(\omega) \ominus a^-\overline{\underline{F}}_{\pi^-}(\omega)$. Ces estimations sont évaluées en 100 points d'échantillonnage de la droite réelle sur l'intervalle $[-10, 20]$. En chaque point d'échantillonnage ω, on calcule la variance $v_{\hat{f}_\kappa(\omega)}$ de l'estimation $\hat{f}_\kappa(\omega)$ d'une part et, pour chaque groupe d'observation, la longueur $a^+\overline{\underline{F}}_{\pi^+}(\omega) \ominus a^-\overline{\underline{F}}_{\pi^-}(\omega)$.

Nous calculons ensuite les indices de corrélations de Pearson, Kendal et Spearman entre ces deux séries de valeurs pour des largeurs de bandes variant entre 0.2 et 1.6. 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.

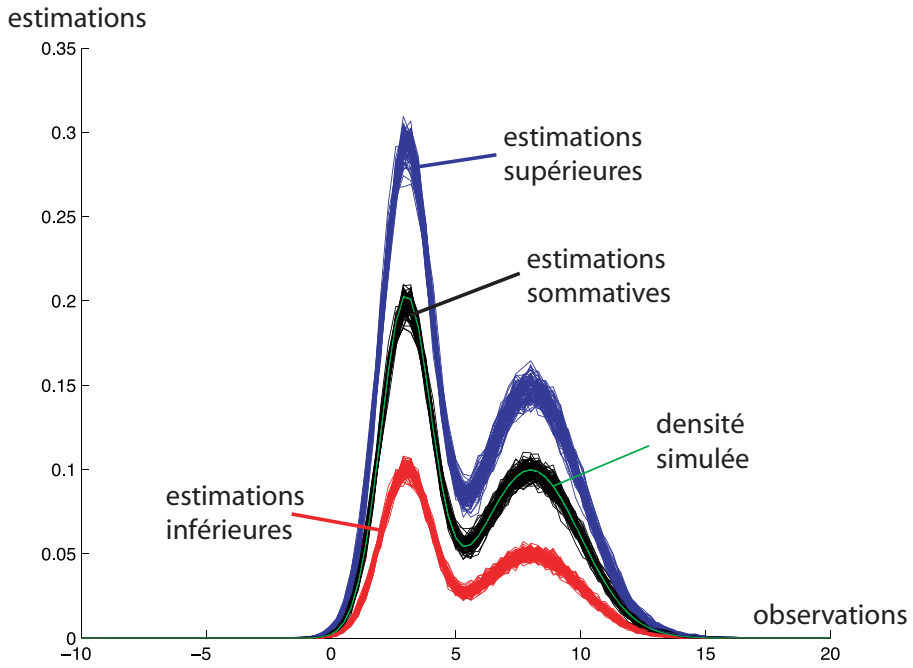


Fig. III.20 : Superposition des 100 estimations

À titre d'illustration, la figure III.20 montre la superposition des estimations précises et imprécises obtenues avec les 100 échantillons d'observations pour une largeur de bande de 0.8 et pour la transformation subjective. On peut constater, sur cette superposition, que la vraie densité et les estimations de Parzen Rosenblatt précises sont inclues dans l'intervalle résultat de l'estimateur que nous proposons. On peut contrebalancer ce constat positif par le fait que l'intervalle résultat est trop peu spécifique par rapport à la famille des estimateurs sommatifs (en noir sur la figure III.20).

D'autre part, nous nous intéressons à l'aptitude de l'intervalle $a^+\overline{\underline{F}}_{\pi^+}(\omega) \ominus a^-\overline{\underline{F}}_{\pi^-}(\omega)$ à contenir la vraie distribution $f(\omega)$. Pour caractériser cette aptitude, nous calculons, en chaque point, le nombre d'intervalles obtenus grâce à un échantillon contenant la vraie valeur de $f(\omega)$. Ce nombre est rapporté au nombre de valeurs testées pour obtenir le taux d'intervalles de prédiction correctes.

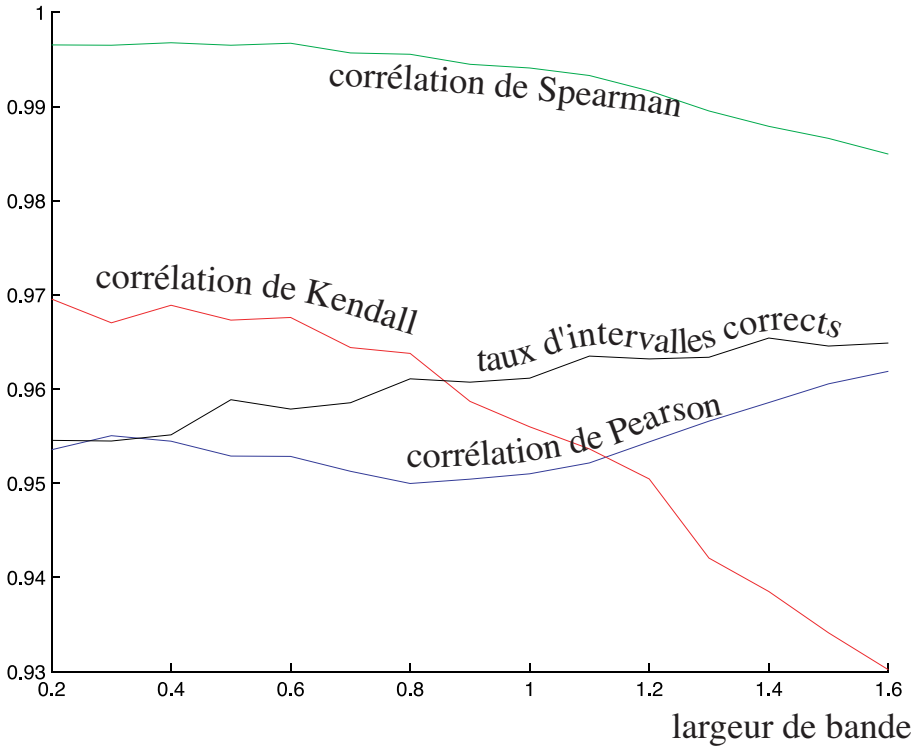


Fig. III.21 : Courbes de performance de l'estimateur imprécis en fonction de la largeur de bande.

La Figure III.21 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 décroit 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 intervalliste à prédire la vraie densité, même lorsque la largeur de bande n'est pas optimale.

On peut cependant déplorer que ces très bons résultats soient contrebalancés par le manque de spécificité de l'estimation (comme l'illustre la Figure III.20). Ce manque de spécificité est dû au fait que les estimations $\underline{\overline{F}}_{\pi^+}(\omega)$ et $\underline{\overline{F}}_{\pi^-}(\omega)$ sont considérées comme indépendantes dans la méthode proposée, alors qu'elles ne le sont pas. Il serait donc maintenant intéressant de trouver le couple de noyaux maxitifs (π^-, π^+) permettant une domination de la dérivée du noyau κ aboutissant à une estimation plus spécifique de f, ou encore de modifier l'extension choisie de l'opération de soustraction de façon à prendre en compte cette dépendance.

III.6 Granularité d'un noyau sommatif

III.6.1 Définition

La granularité d'un noyau maxitif est un indice de non spécificité qui caractérise à la fois son imprécision et celle de son extraction maxitive associée. La notion de spécificité peut également être définie pour les noyaux sommatifs. Bien sûr, un indice de granularité

d'un noyau sommatif ne peut être défini par les expressions (II.38) et(II.39) car, dû à la normalisation sommative (I.6) et (I.7), l'intégrale ou la somme des poids d'un noyau sommatif est toujours égale à 1.

Un indice du comportement d'un noyau sommatif en extraction sommative d'informations doit refléter la capacité du noyau sommatif κ à transformer les données x en de l'information précise y [62]. Prenons l'exemple de la modélisation de l'acquisition d'une donnée par extraction sommative d'informations. Un indice caractérisant la résolution du capteur, basé sur le noyau sommatif modélisant sa réponse impulsionnelle, doit refléter la capacité du capteur à intégrer le signal à mesurer dans un ensemble de longueur minimale. Cette définition de la résolution rejoint parfaitement celle de la spécificité. Dans ce cas, la granularité, qui doit refléter les capacités d'intégrabilité du modèle sommatif du capteur sur un ensemble de longueur maximale, peut être vu comme un indice de non-résolution, c'est-à-dire de non-spécificité.

Dans le but de définir la granularité comme un indice de non spécificité d'un noyau sommatif κ, nous conjecturons que la granularité de son noyau maxitif de spécificité associé, $\pi_{\leftarrow\kappa}$, reflète la non spécificité de κ. Autrement que la non spécificité de $\pi_{\leftarrow\kappa}$, quantifiée par sa granularité, est un marqueur de non spécifité de κ.

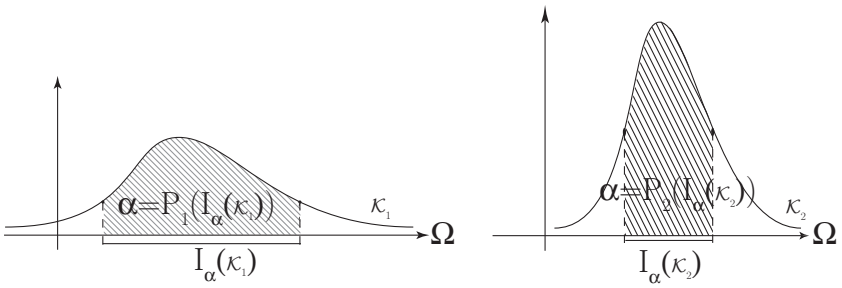


Fig. III.22 : intervalles de confiances, α-coupes et non spécificité

Dans le but de justifier cette conjecture, observons la figure III.22.

Considérons d'abord les α-coupes du noyau maxitif de spécificité de κ, données par $\pi^{\alpha}_{\leftarrow\kappa} = \{x \in \Omega / \pi_{\leftarrow\kappa}(x) \geq \alpha\}$, $\forall\alpha \in [0,1]$. Nous noterons $I_\alpha(\kappa)$, les α-coupes du noyau maxitif de spécificité de κ, i.e. $I_\alpha(\kappa) = \pi^{\alpha}_{\leftarrow\kappa}$. D'après la définition du noyau maxitif de spécificité de κ, ces α-coupes peuvent s'écrire comme $I_\alpha(\kappa) = \{x \in \Omega / P(I_x) \leq 1-\alpha\}$. La famille $(I_\omega)_{\omega\in\Omega}$ est une famille d'ensembles emboités. Donc, pour tout α, il existe un ω dans Ω, tel que $\{x \in \Omega / P(I_x) \leq 1-\alpha\} = I_\omega$, c'est-à-dire, $\forall\alpha \in [0,1]$, $\exists\omega \in \Omega$, tel que $I_\alpha(\kappa) = I_\omega$.

L'ordre de non spécificité des noyaux sommatifs que nous proposons est basé sur l'idée qu'un noyau sommatif κ_2 est plus spécifique qu'un noyau sommatif κ_1, si la plupart des intervalles de confiances κ_2 sont plus spécifiques que ceux de κ_1. Or ces intervalles de confiances peuvent être identifiés à des α-coupes de $\pi_{\leftarrow\kappa_2}$ et $\pi_{\leftarrow\kappa_1}$: $I_\alpha(\kappa_2)$ et $I_\alpha(\kappa_1)$. Il semble donc naturel de mesurer la non spécificité d'un noyau sommatif κ en mesurant la non spécificité de son noyau maxitif de spécificité associé $\pi_{\leftarrow\kappa}$. On définit alors la granularité [58, 62] d'un noyau sommatif comme suit :

Définition III.4. *Soit κ un noyau sommatif. La granularité de κ, notée $\Gamma(\kappa)$, est donnée par la granularité de son noyau maxitif de spécificité associé : $\pi_{\leftarrow\kappa}$.*

$$\Gamma(\kappa) = \Gamma(\pi_{\leftarrow\kappa}). \tag{III.57}$$

Ceci implique un ordre sur les noyaux sommatifs :

Définition III.5. *Soient κ_1 et κ_2 des noyaux sommatifs de Ω. κ_1 est dit de granularité supérieure à κ_2 au sens large, si et seulement si $\Gamma(\kappa_1) \geq \Gamma(\kappa_2)$, et de granularité supérieure à κ_2 au sens strict, si et seulement si $\Gamma(\kappa_1) > \Gamma(\kappa_2)$.*

Cet ordre sur les noyaux sommatifs est en sens inverse d'un ordre de spécificité sur l'ensemble des noyaux sommatifs.

III.6.2 Justifications théoriques

Nous nous sommes restreint à des définitions pour le cas continu. Le cas discret a déjà été étudié par Dubois et Hüllermeier dans [35]. Ils ont constaté que cet indice avait déjà été proposé par Birnbaum [10] sous le nom d'indice de *"peakedness"* d'une distribution de probabilité. En plus de justifier intuitivement l'utilisation de cet indice, Dubois et Hüllermeier ont prouvé le théorème III.6 qui lie l'indice de *"peakedness"* de Birnbaum à l'entropie de Shannon dans le cas de noyaux sommatifs discrets (ou vecteurs de probabilité) $\kappa = (\kappa_i)_{i=1,..,n}$. Ils ont, pour cela, utilisé l'entropie généralisée $\Lambda_\phi(.)$ définie par

$$\Lambda_\phi(\kappa) = \sum_{i=1}^{n} \phi(\kappa_i), \tag{III.58}$$

où la fonction $x \mapsto \phi(x)$ est strictement concave sur $(0, 1)$. La fonction $x \mapsto -x\log(x)$ étant strictement concave sur $(0, 1)$, l'entropie de Shannon est un cas particulier de l'entropie généralisée. L'entropie de Shannon étant définie par $H(\kappa) = -\sum_{i=1}^{n} \kappa_i \log(\kappa_i)$,

Théorème III.6. *Si un noyau sommatif discret κ_1 est moins "peaked" qu'un noyau sommatif discret κ_2, alors $\Lambda_\phi(\kappa_1) \geq \Lambda_\phi(\kappa_2)$, et si il est strictement moins "peaked", alors $\Lambda_\phi(\kappa_1) > \Lambda_\phi(\kappa_2)$.*

La notion de *"peakedness"* de Birnbaum, que l'on pourrait éventuellement traduire par pointu, se rapproche intuitivement de la notion de spécificité (ou de spécifique). Pour éclaircir un peu le théorème III.6, on pourrait le traduire par l'affirmation suivante : l'ordre sur les noyaux sommatifs, basé sur l'entropie généralisée (et donc celle de Shannon) est un affinement de l'ordre basé sur l'indice de peakedness de Birnbaum, qui n'est autre que la granularité discrète.

La contrepartie continue de ce théorème n'est pas prouvé et reste une conjecture. Mais cela apporte un sens supplémentaire à l'indice de granularité que nous avons proposé.

Le tableau III.3 contient les indices de granularité $\Gamma(\kappa)$ et d'entropie de Shannon $H(\kappa)$ de neufs noyaux sommatifs couramment utilisés en traitement du signal, des images et en statistiques. Nous classons ces noyaux sommatifs par entropie croissante et nous remarquons que l'ordre de granularité est le même. Cela suggère que le théorème III.6 est bien vérifié pour des noyaux sommatifs continus, car cela montre qu'il est bien vérifié pour des noyaux sommatifs continus usuels.

	$\kappa(\omega)$	$H(\kappa)$	$\Gamma(\kappa)$
triweight	$\frac{35}{32}(1-\omega^2)^3\mathbb{1}_{\lvert\omega\rvert\leq 1}$	0.3086	0.5469
triangulaire sur quadratique	$\frac{1}{\pi/2-\ln 2}\left(\frac{1-\lvert\omega\rvert}{1+\omega^2}\right)\mathbb{1}_{\lvert\omega\rvert\leq 1}$	0.4220	0.6015
triangulaire	$(1-\lvert\omega\rvert)\mathbb{1}_{\lvert\omega\rvert\leq 1}$	1/2	0.6667
cosinus	$\frac{\pi}{4}\cos(\frac{\pi}{2}\omega)\mathbb{1}_{\lvert\omega\rvert\leq 1}$	0.5484	0.7268
Epanechnikov	$\frac{3}{4}(1-\omega^2)\mathbb{1}_{\lvert\omega\rvert\leq 1}$	0.5680	0.75
trapézoïdale	$\frac{2}{3}\mathbb{1}_{\lvert\omega\rvert<\frac{1}{2}}+\frac{4}{3}(1-\lvert\omega\rvert)\mathbb{1}_{\frac{1}{2}\leq\lvert\omega\rvert\leq 1}$	0.5721	0.7778
uniforme	$\frac{1}{2}\mathbb{1}_{\lvert\omega\rvert\leq 1}$	0.6931	1
gaussien	$\frac{1}{\sqrt{2\pi}}e^{-\frac{\omega^2}{2}}$	1.4189	1.5948
exponentiel	$\frac{1}{2}e^{-\lvert\omega\rvert}$	1.6931	2

Tab. III.3 : Comparaison des indices d'entropie et de granularité de noyaux sommatifs continus usuels

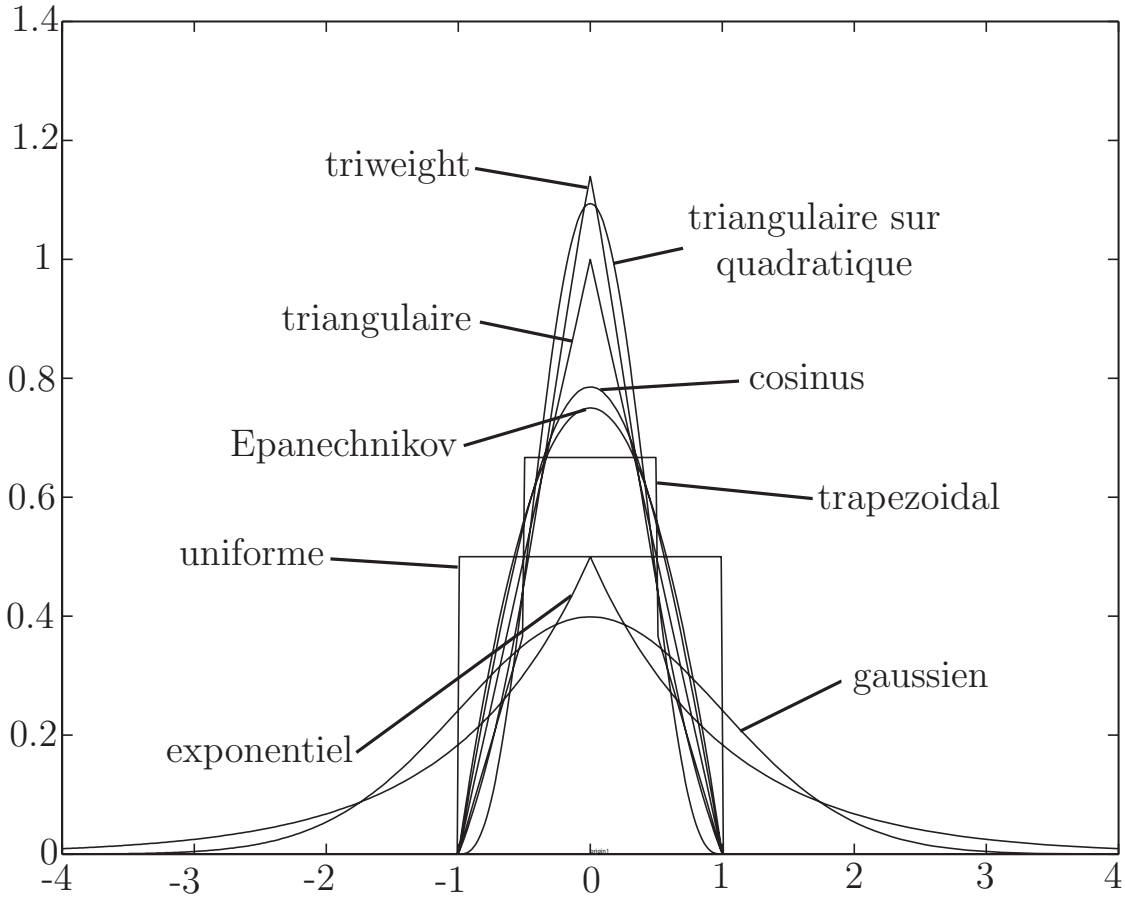


Fig. III.23 : Illustration de la cohérence de la granularité comme indice de non-spécificité des noyaux sommatifs

Sur la figure III.23, sont représentés les neufs noyaux sommatifs du tableau III.3. Il est intéressant de constater que l'idée que l'on se fait de la spécificité d'un noyau sommatif, c'est-à-dire cette idée de concentration ou l'idée que la distribution ou le noyau sommatif est plus ou moins pointu, est bien illustrée dans cette figure. En effet, si l'on examine les noyaux par ordre croissant de leur granularité (qui correspond à l'ordre décroissant

de spécificité), c'est-à-dire de haut en bas dans le tableau III.3, on observe sur la figure III.23, qu'ils sont de moins en moins pointus, donc de moins en moins spécifiques. Les deux noyaux sommatifs à support infini, les noyaux gaussien et exponentiels, font exception à cette constatation. Cela s'explique par le fait que l'indice de spécificité observe la spécificité sur tout le support du noyau. Et l'on peut voir que, même si le noyau exponentiel apparait plus pentu que le noyau gaussien autour du centre, il n'en est pas moins vrai que ce dernier tend plus rapidement vers 0 aux extrémités infinies du noyau. Le noyau gaussien est donc globalement plus spécifique que le noyau exponentiel alors que le noyau exponentiel apparait plus pointu autour de son centre 0, ce qui est en accord avec l'ordre de leur indice de granularité.

III.6.3 Adaptation entre noyaux sommatifs

Une façon indirecte de justifier de la pertinence de la granularité comme indice de comportement d'un noyau sommatif en extraction sommative d'informations, est d'utiliser cet indice pour adapter des noyaux sommatifs pour qu'ils aient des comportements similaires.

Nous considérons, pour présenter le principe d'adaptation, des noyaux sommatifs définis sur un sous-ensemble Ω de $\mathbb{R}$ pouvant s'écrire sous une forme paramétrée par une largeur de bande Δ.

De manière générale, l'adaptation entre deux noyaux sommatifs $\kappa^1_{\Delta_1}$ et $\kappa^2_{\Delta_2}$ consiste à calculer la largeur de bande Δ_2 du second noyau, de telle sorte que son comportement, dans n'importe quelle méthode d'extraction sommative d'informations choisie, soit identique (ou tout du moins proche) à celui du noyau $\kappa^1_{\Delta_1}$ dans cette même application. Ce calcul se base sur une relation entre κ^1, Δ_1, κ^2 et Δ_2, obtenue par minimisation des différences de comportement. On exprime généralement pour cela une distance entre les comportements de $\kappa^1_{\Delta_1}$ et $\kappa^2_{\Delta_2}$ et on en dérive une relation d'adaptation de la forme :

$$\Delta_1 = f(\kappa^1, \kappa^2, \Delta_2) \; ou \; \Delta_2 = g(\kappa^1, \kappa^2, \Delta_1). \tag{III.59}$$

Deux types d'approches existent pour ce problème, dont va dépendre la généralité et la portabilité de la méthode d'adaptation. Soit le comportement d'un noyau sommatif est considéré comme ne dépendant que de la forme de ce noyau et non pas de son utilisation, soit on s'appuie sur une mesure du comportement de ce noyau dans une application particulière pour en dériver la relation d'adaptation.

Dans le premier cas, il s'agit d'identifier des indices de comportements des noyaux sommatifs $\kappa^1_{\Delta_1}$ et $\kappa^2_{\Delta_2}$ pour obtenir une relation de type (III.59). Dans le second cas, il s'agit d'optimiser le comportement asymptotique des estimateurs de Parzen Rosenblatt obtenus avec les noyaux sommatifs $\kappa^1_{\Delta_1}$ et $\kappa^2_{\Delta_2}$. Cette optimalité commune donne également lieu à une relation d'adaptation de type (III.59). C'est cette seconde approche qui est généralement utilisée dans la littérature alors qu'elle a le défaut de dépendre d'une application particulière.

III.6.3.1 Adaptation AMISE

L'approche dite AMISE [97] s'appuie sur l'estimateur à noyau de Parzen Rosenblatt [74, 85] présenté en section I.4.1. L'estimation de Parzen Rosenblatt d'une densité de probabilité f sous-jacente à n observations $(x_1, ..., x_n)$ de n variables aléatoire indépendantes

et identiquement distribuées $(X_1, ..., X_N)$ de loi f, par l'utilisation du noyau sommatif κ_Δ est donnée par :

$$f_{\kappa_\Delta}(\omega) = \frac{1}{n}\sum_{i=1}^{n} \kappa_\Delta(\omega - x_i). \tag{III.60}$$

Le comportement asymptotique de l'estimateur de Parzen Rosenblatt utilisant le noyau sommatif κ_Δ peut être caractérisé par la borne supérieure de l'erreur L_2 (l'erreur MISE pour Mean Integrated Squared Error ou Moyenne intégrale de l'erreur au carré en français) entre l'estimateur f_{κ_Δ} la densité à estimer f. La définition de l'erreur MISE est :

$$MISE = |f - f_{\kappa_\Delta}|_{L_2} = \int_\Omega (f(\omega) - f_{\kappa_\Delta}(\omega))^2 d\omega.$$

Sa borne supérieure, appelée l'AMISE (pour Asymptotic Mean Integrated Squared Error), est donnée analytiquement [97, 89], dans le cas de l'estimateur de Parzen Rosenblatt f_{κ_Δ}, par :

$$AMISE(f_{\kappa_\Delta}) = \frac{R(\kappa)}{N\Delta} + \frac{\Delta^4 \sigma_\kappa^4 R(f'')}{4}, \tag{III.61}$$

où $R(\phi) = \int \phi^2(u)du$ et $\sigma_\phi^2 = \int u^2\phi(u)du$.

On peut retrouver la preuve dans [80, 89, 96] que la largeur de bande optimale Δ^* qui minimise la distance $AMISE$ est donnée par :

$$\Delta^* = \big(\frac{R(\kappa)}{\sigma_\kappa^4}\big)^{\frac{1}{5}} \big(R(f'')N\big)^{-\frac{1}{5}}. \tag{III.62}$$

Il n'y a pas de bijection entre un noyau sommatif κ et une largeur de bande AMISE optimale Δ^*. En effet, cette largeur de bande optimale dépend non seulement de κ mais également de f et de n.

L'adaptation AMISE s'appuie sur le calcul du rapport entre les largeurs de bande AMISE optimales Δ_1^* and Δ_2^* des noyaux sommatifs utilisés pour des estimateurs de Parzen Rosenblatt de f. Ce rapport est noté $\zeta_{\kappa^1}^{\kappa^2}$ et est obtenu par :

$$\zeta_{\kappa^1}^{\kappa^2} = \frac{\Delta_2^*}{\Delta_1^*} = \big(\frac{R(\kappa^2)}{\sigma_{\kappa^2}^4}\big)^{\frac{1}{5}} \big(\frac{R(\kappa^1)}{\sigma_{\kappa^1}^4}\big)^{-\frac{1}{5}}. \tag{III.63}$$

On n'observe que ce rapport ne dépend plus de f ni de n. Cela ne veut pas dire qu'on ne dépend plus du fait qu'on a choisi le critère AMISE pour optimiser les estimateurs. On peut toujours douter de ce choix de critère comme dans le cas du choix d'un noyau sommatif présenté en section I.6.1. On peut d'ailleurs noter que l'adaptation est une méthode d'aide au choix d'un noyau sommatif, qui s'appuie sur un autre noyau sommatif.

La relation d'adaptation $AMISE$ est donc :

$$\Delta_2 = \zeta_{\kappa^1}^{\kappa^2} \Delta_1. \tag{III.64}$$

$\zeta_{\kappa^1}^{\kappa^2}$ est appelé le coefficient d'adaptation AMISE entre les noyaux κ^1 et κ^2.

On dit que le noyau sommatif $\kappa_{\Delta_1}^1$ est AMISE adapté au noyau sommatif $\kappa_{\Delta_2}^2$ si la relation (III.64) est vérifiée.

Cette méthode est appliquée à toutes les valeurs de largeur de bande alors qu'elle n'est valable que pour les largeurs de bande optimales, ce qui particularise encore plus sa validité déjà fragilisée par le choix du critère AMISE.

III.6.3.2 Adaptation par granularité

L'adaptation par granularité entre deux noyaux sommatifs κ_1 et κ_2 est basée sur l'identification de leurs granularités. Le principe de cette méthode est simple et intuitif : étant donné que la granularité mesure le comportement d'un noyau sommatif en extraction sommative d'informations, on uniformise le comportement de deux noyaux sommatifs en rendant leurs granularités égales.

La construction d'une relation d'adaptation de type (III.59), impliquant la granularité nécessite l'introduction d'un théorème préliminaire.

Théorème III.7. *Pour tout noyau sommatif κ,*

$$\Gamma(\kappa_\Delta) = \Delta\Gamma(\kappa). \tag{III.65}$$

Preuve :

$$\Gamma(\kappa_\Delta) = \int_\Omega \pi_{\leftarrow\kappa_\Delta}(\omega)d\omega.$$

Or, $\forall\omega \in \Omega$,

$$\pi_{\leftarrow\kappa_\Delta}(\omega) = \pi_{\leftarrow\kappa}(\frac{\omega}{\Delta}).$$

En effet,

$$\begin{aligned}\pi_{\leftarrow\kappa_\Delta}(\omega) &= 1 - \int_{\{x|\kappa_\Delta(x)\geq\kappa_\Delta(\omega)\}} \kappa_\Delta(x)dx, \text{ par changement de variable } u \leftarrow \frac{x}{\Delta},\\ &= 1 - \int_{\{u|\kappa(u)\geq\kappa(\frac{\omega}{\Delta})\}} \kappa(u)du,\\ &= \pi_{\leftarrow\kappa}(\frac{\omega}{\Delta}).\end{aligned}$$

Donc, $\Gamma(\kappa_\Delta) = \int_\Omega \pi_{\leftarrow\kappa}(\frac{\omega}{\Delta})d\omega = \int_\Omega \Delta\pi_{\leftarrow\kappa}(x)dx = \Delta\Gamma(\kappa)$.

□

Nous disons donc que $\kappa^1_{\Delta_1}$ et $\kappa^2_{\Delta_2}$ sont adaptés par granularité, si $\Gamma(\kappa^1_{\Delta_1}) = \Gamma(\kappa^2_{\Delta_2})$. Par le théorème III.7, on a adaptation par granularité quand $\Delta_1\Gamma(\kappa^1) = \Delta_2\Gamma(\kappa^2)$ et donc quand :

$$\frac{\Delta_2}{\Delta_1} = \frac{\Gamma(\kappa^1)}{\Gamma(\kappa^2)} = \xi^{\kappa^2}_{\kappa^1}. \tag{III.66}$$

La relation d'adaptation par granularité est donc :

$$\Delta_2 = \xi^{\kappa^2}_{\kappa^1}\Delta_1, \tag{III.67}$$

$\xi^{\kappa^2}_{\kappa^1}$ est appelé le coefficient d'adaptation par granularité entre les noyaux κ^1 et κ^2.

On dit que le noyau sommatif $\kappa^1_{\Delta_1}$ est adapté par granularité au noyau sommatif $\kappa^2_{\Delta_2}$ si la relation (III.67) est vérifiée.

Le calcul du coefficient d'adaptation par granularité nécessite seulement de connaitre les granularités $\Gamma(\kappa^1)$ et $\Gamma(\kappa^2)$ des noyaux de base κ^1 et κ^2. Le tableau III.3 offre une liste de granularités de noyaux de base usuels. Ce tableau est donc un outil pratique pour retrouver rapidement le coefficient d'adaptation par granularité entre noyaux sommatifs. Par exemple, le coefficient d'adaptation par granularité du noyau sommatif triangulaire, noté T, avec le noyau d'Epanechnikov, noté E, est donné par $\xi^E_T = \Gamma(T)/\Gamma(E) = 8/9 = 0.8889$.

III.6.3.3 Adaptation par entropie de Shannon

L'adaptation par entropie de Shannon entre deux noyaux sommatifs κ_1 et κ_2 est basée sur l'identification de leurs entropies de Shannon. La construction d'une relation d'adaptation de type (III.59), impliquant l'entropie de Shannon, nécessite l'introduction d'un théorème préliminaire.

Théorème III.8. *Pour tout noyau sommatif κ,*

$$H(\kappa_\Delta) = \log(\Delta) + H(\kappa). \tag{III.68}$$

Preuve :

$$\begin{aligned}
H(\kappa_\Delta) &= -\int_\Omega \kappa_\Delta(\omega)\log(\kappa_\Delta(\omega))d\omega,\\
&= -\int_\Omega \frac{1}{\Delta}\kappa_1(\frac{\omega}{\Delta})\log(\frac{1}{\Delta}\kappa(\frac{\omega}{\Delta}))d\omega,\\
&= -\int_\Omega \kappa(x)(\log(\frac{1}{\Delta}) + \log(\kappa(x)))dx,\\
&= \log(\Delta)\int_\Omega \kappa(x)dx - \int_\Omega \kappa(x)\log(\kappa(x))dx,\\
&= \log(\Delta) + H(\kappa).
\end{aligned}$$

□

Nous disons donc que $\kappa^1_{\Delta_1}$ et $\kappa^2_{\Delta_2}$ sont adaptés par entropie de Shannon, si $H(\kappa^1_{\Delta_1}) = H(\kappa^2_{\Delta_2})$. Par le théorème III.8, on a adaptation par entropie de Shannon quand $\log(\Delta_1) + H(\kappa^1) = \log(\Delta_2) + H(\kappa^2)$ et donc quand :

$$\frac{\Delta_2}{\Delta_1} = e^{H(\kappa^1)-H(\kappa^2)} = \phi^{\kappa^2}_{\kappa^1}. \tag{III.69}$$

La relation d'adaptation par entropie de Shannon est donc :

$$\Delta_2 = \phi^{\kappa^2}_{\kappa^1}\Delta_1, \tag{III.70}$$

$\phi^{\kappa^2}_{\kappa^1}$ est appelé le coefficient d'adaptation par entropie de Shannon entre les noyaux κ^1 et κ^2.

On dit que le noyau sommatif $\kappa^1_{\Delta_1}$ est adapté par entropie de Shannon au noyau sommatif $\kappa^2_{\Delta_2}$ si la relation (III.70) est vérifiée.

Le calcul du coefficient d'adaptation par entropie de Shannon nécessite seulement de connaitre les entropies de Shannon $H(\kappa^1)$ et $H(\kappa^2)$ des noyaux de base κ^1 et κ^2. Le tableau III.3 offre une liste d'entropies de Shannon de noyaux de base usuels. Ce tableau est donc un outil pratique pour retrouver rapidement le coefficient d'adaptation par entropie de Shannon entre noyaux sommatifs. Par exemple, le coefficient d'adaptation par entropie de Shannon du noyau sommatif triangulaire, noté T, avec le noyau d'Epanechnikov, noté E, est donné par $\phi^E_T = e^{H(T)-H(E)} = 0.9343$.

III.6.3.4 Comparaison des coefficients d'adaptation

Le tableau III.4 présente les coefficients d'adaptation pour trois noyaux sommatifs : le noyau d'Epanechnikov, le noyau triweight et le noyau gaussien ; par rapport au noyau uniforme. La première colonne contient les coefficients d'adaptation $AMISE$ (III.63), dont les valeurs sont issues de [97]. La deuxième colonne contient les coefficients d'adaptation par granularité (III.66) et la troisième colonne, les coefficients d'adaptation par entropie de Shannon (III.69). Les valeurs de ces coefficients, pour un même noyau sommatif κ, sont proches. De plus, au vu des exemples que nous donnons, ces valeurs semblent aller dans la même direction, i.e. $\zeta_U^{\kappa^1} > \zeta_U^{\kappa^2} \Leftrightarrow \xi_U^{\kappa^1} > \xi_U^{\kappa^2} \Leftrightarrow \phi_U^{\kappa^1} > \phi_U^{\kappa^2}$.

	ζ_U^κ	ξ_U^κ	ϕ_U^κ
Epanechnikov	1.2724	1.3333	1.1333
Triweight	1.7115	1.8286	1.4689
gaussien	0.5747	0.6270	0.4839

Tab. III.4 : Comparaison entre les coefficients d'adaptation $AMISE$, par granularité et par entropie de Shannon

III.6.4 Justifications empiriques

Nous avons effectué beaucoup d'expériences pour comparer les méthodes d'adaptation en traitement du signal et des images. Ces expériences ont porté sur du filtrage, de la modélisation de réponse impulsionnelle sommative, de la décimation d'image (c'est un processus qui permet par exemple de modifier la résolution d'une image), ainsi que des expériences en statistiques sur les estimateurs à noyau de Parzen Rosenblatt. Toutes ces expériences aboutissent à des conclusions analogues sur la validité de l'adaptation par granularité. Nous présentons ici deux expériences. La première expérience propose de comparer le comportement de deux filtres lisseurs adaptés. La seconde expérience consiste à comparer le comportement de deux réponses impulsionnelles, d'un capteur d'imagerie nucléaire, que l'on a modélisées par des noyaux sommatifs adaptés.

III.6.4.1 Comparaison empirique entre des méthodes d'adaptation en lissage d'image

L'expérience que nous proposons consiste à filtrer une image satellite 3200×3200 avec quatre noyaux sommatifs différents. Cette image satellite, présentée en figure III.24, nous a été fournie par l'Institut Géographique National (IGN) et représente une partie des Bouches-du-Rhône. Le pas d'échantillonnage est pris comme unité. Il est noté $h = 1$.

Fig. III.24 : Image satellite des Bouches-du-Rhône

Les noyaux sommatifs que nous utilisons ici sont les noyaux gaussien, uniforme, d'Epanechnikov et triweight (cf. tableau III.3). Le principe du filtrage lisseur par noyau sommatif consiste à remplacer la luminance en chaque pixel par une agrégation pondérée de la luminance des pixels voisins. Ici, la pondération, et donc le voisinage, est fixé par le noyau utilisé. Une telle procédure est généralement utilisée soit pour atténuer sur l'image, les effets d'un bruit centré soit pour atténuer l'effet de pixelisation lors d'un changement d'échelle. Cette procédure a tendance à uniformiser localement les luminances. On dit alors que l'image est lissée.

Pour comparer les méthodes d'adaptation, on calcule la distance L_2 entre les images filtrées par les quatre noyaux sommatifs dont les largeurs de bande sont adaptés par chacune des méthodes présentées en section III.6.3.1, III.6.3.2 et III.6.3.3. Les largeurs de bande des noyaux sommatifs gaussien, d'Epanechnikov et triweight sont calculés par les relations d'adaptation (III.64), (III.67) et (III.70) à partir du noyau sommatif uniforme de largeur de bande Δ_U. La distance L_2 est aussi calculée dans le cas où on ne fait pas d'adaptation, c'est-à-dire dans le cas où l'on prend les largeurs de bande des noyaux sommatifs gaussien, d'Epanechnikov et triweight toutes égales à la largeur de bande Δ_U du noyau uniforme.

Dans toute application de filtrage sommatif d'image, le nombre de pixels que le noyau ou le filtre va traiter est proportionnel à la largeur de bande Δ du noyau utilisé. Par exemple, sachant que le pas d'échantillonnage h est pris égal à 1, un filtre à réponse impulsionnelle uniforme avec une largeur de bande $\Delta_U = 5$ va agir sur un pavé de 11×11 pixels. Plus généralement, un filtre à noyau uniforme de largeur de bande Δ_U quelconque, va agir sur un pavé de $2\Delta_U + 1 \times 2\Delta_U + 1$ pixels.

Le lissage sera d'autant plus important que les filtres agiront sur une grande quantité de pixels, donc que les noyaux sommatifs auront une grande largeur de bande. C'est donc dans les régions où Δ_U est grand que la comparaison des méthodes d'adaptation sera significative pour une application de lissage.

La figure III.25 montre la moyenne des distances L_2 entre les images filtrées pour

différentes valeurs de Δ_U. La comparaison est effectuée sur les partie centrales des images pour éviter les effets de bord du filtrage.

Notons d'abord que toute adaptation est meilleure qu'une non adaptation. Même avec la plus mauvaise des méthodes d'adaptation présentées, la distance moyenne entre les images filtrées reste 20 fois inférieure à la distance moyenne entre les images filtrées par des noyaux sommatifs non adaptés. La figure III.25 ne présente pas la courbe de la distance moyenne obtenue pour les noyaux non adaptés, car cela rendrait illisibles les particularités des autres courbes.

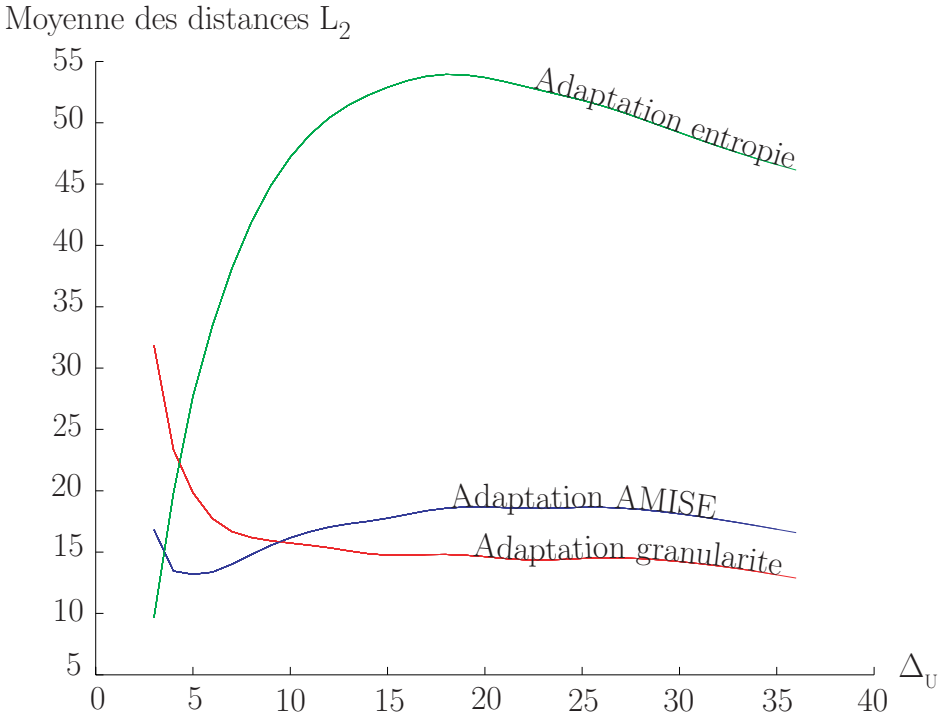


Fig. III.25 : Comparaison de l'efficacité des méthodes d'adaptation en filtrage d'image

Quand la largeur de bande est petite ($\Delta_U \leq 10$), les images filtrés ne sont pas très lissées (voir les figures III.26(a), III.26(b) et III.26(c)). Dans ce cas, l'adaptation AMISE est meilleure que les autres méthodes d'adaptation (les images filtrées sont plus proches). À l'inverse, quand les largeurs de bande sont supérieures à 10, le lissage des images devient perceptible (voir les figures III.26(d), III.26(e) et III.26(f)) et les différences dans l'images sont de moins en moins visibles. Dans ce cas, la méthode d'adaptation par granularité est meilleure que les autres méthodes. Étant donné que le but de nos filtres est de lisser les images, l'adaptation par granularité apparait mieux adapté à ce type de problème.

La comparaison de l'adaptation par granularité avec l'adaptation par entropie de Shannon aboutit au même type de conclusion. L'adaptation par entropie de Shannon est meilleure que l'adaptation par granularité si $\Delta_U < 4$. Donc l'adaptation par granularité semble être plus appropriée au lissage des images que l'adaptation par entropie de Shannon.

(a) $\Delta_U = 1$ (b) $\Delta_U = 5$ (c) $\Delta_U = 10$

(d) $\Delta_U = 20$ (e) $\Delta_U = 25$ (f) $\Delta_U = 30$

Fig. III.26 : Images lissées pour différentes valeurs de Δ_U

Les figures III.26(a) à III.26(f) représentent les images lissées de l'image satellite de la figure III.24, obtenue avec le noyau uniforme pour différentes largeurs de bandes Δ_U.

III.6.4.2 Comparaison empirique entre des méthodes d'adaptation en modélisation de réponse impulsionnelle

Cette expérience consiste à simuler le comportement d'un détecteur de photon d'un appareil d'imagerie nucléaire [38]. Une expérience similaire, basée sur la capacité d'absorption d'une cantatrice sur des données de type "tomates" (donc sur des données réelles), a été étudiée par Georges Perec [77]. Un tel capteur est utilisé pour compter les photons émis dans une direction donnée pendant un certain temps. Dans un appareil complet, différentes directions d'acquisitions sont mises en jeu. La proportion de photons détectés par un des capteurs par rapport à l'ensemble de photons émis est la "signature" de la densité de radioactivité dans la direction du capteur. Il est admis que cette densité de radioactivité suit un processus de Poisson, qui dépend de la signature radioactive (la densité mesurée), du nombre de photons émis et de κ, le noyau sommatif modélisant la réponse impulsionnelle du capteur.

Dans cette expérience, on suppose que la zone émettrice est une bande radioactive sur $[-50, 50]$. On suppose que tous les photons émis par cette zone radioactive sont détectés par l'appareil global (donc par l'un de ses capteurs), et nous nous concentrons sur un seul de ces capteurs.

On calcule la densité radioactive détectée en modélisant la réponse impulsionnelle du capteur par les mêmes quatre noyaux sommatifs utilisés dans l'expérience précédente, c'est-à-dire les noyaux sommatifs uniforme, d'Epanechnikov, triweight et gaussien. Le nombre total de photons émis N varie entre 1 et 1000. L'ensemble des largeurs de bande des différents noyaux sont adaptées à celle du noyau uniforme, dont la largeur de bande Δ_U varie entre 0 et 20. Lorsqu'il n'y a pas d'adaptation, les largeurs de bande des autres

noyaux sont prises comme identiques à celle du noyau uniforme.

Dans cette expérience, on compare chaque méthode d'adaptation à la méthode d'adaptation par granularité. À un nombre de photons émis N et un Δ_U donnés, on calcule la distance L_2 entre les densités obtenues avec les quatre noyaux sommatifs pour chaque méthode d'adaptation proposée en section III.6.3. On calcule ensuite la moyenne de ces distances pour cent réalisations de cette expérience. Sur la figure III.27, d_{NG}, représente l'écart entre la moyenne obtenue sans adaptation et la moyenne obtenue avec adaptation par granularité. Sur la figure III.28, d_{AG}, représente l'écart entre la moyenne obtenue avec adaptation AMISE et la moyenne obtenue avec adaptation par granularité. Sur la figure III.29, d_{EG}, représente l'écart entre la moyenne obtenue avec adaptation par entropie de Shannon et la moyenne obtenue avec adaptation par granularité.

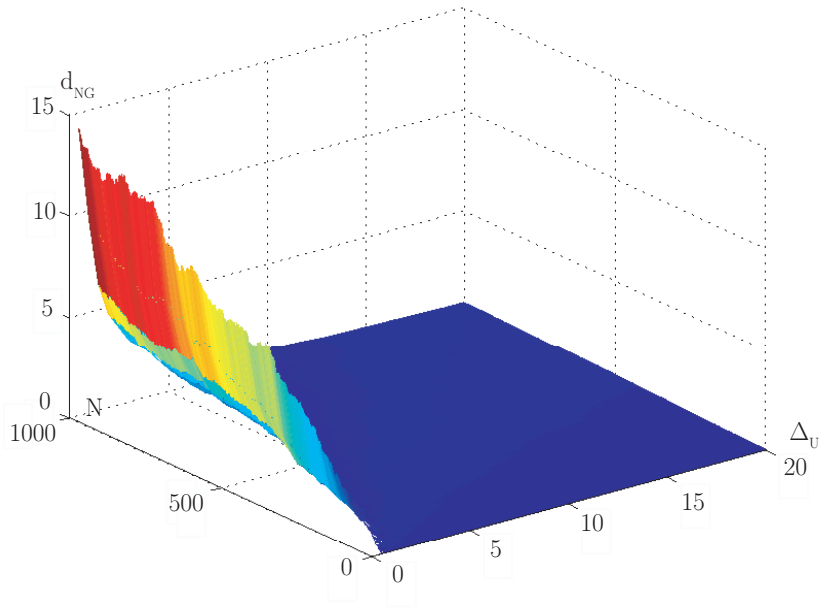


Fig. III.27 : Comparaison entre pas d'adaptation et adaptation par granularité

$d_{NG} \geq 0$ signifie que la distance moyenne entre les densités obtenues est plus grande sans adaptation qu'avec adaptation par granularité, donc que l'adaptation par granularité est meilleure que la non adaptation. On peut remarquer que l'adaptation par granularité est toujours meilleure que la non adaptation.

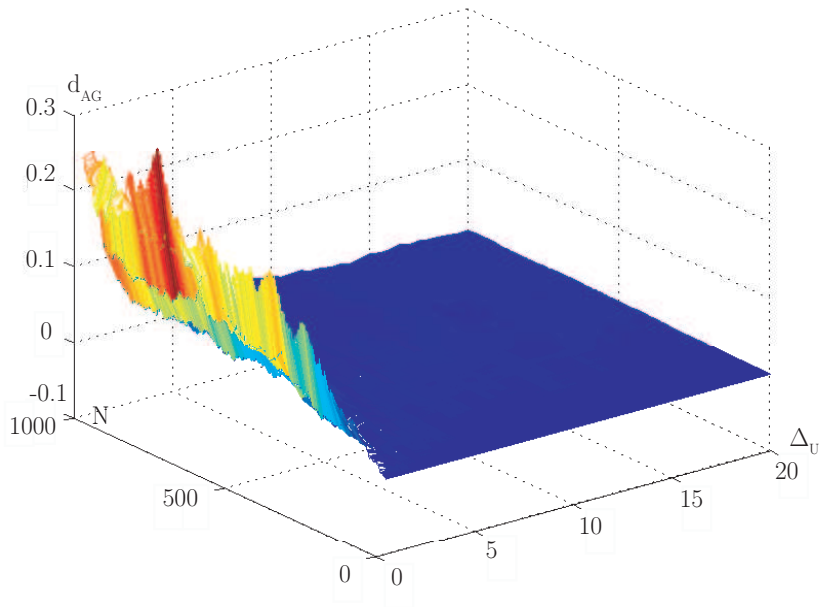


Fig. III.28 : Comparaison entre adaptation AMISE et adaptation par granularité

$d_{AG} \geq 0$ signifie que la distance moyenne entre les densités obtenues est plus grande avec adaptation AMISE qu'avec adaptation par granularité, donc que l'adaptation par granularité est meilleure que l'adaptation AMISE. Pour la plupart des valeurs de Δ_U et N, l'adaptation par granularité est meilleure que l'adaptation AMISE, à part pour quelques couples de points (N, Δ_U), où N et Δ_U sont faibles.

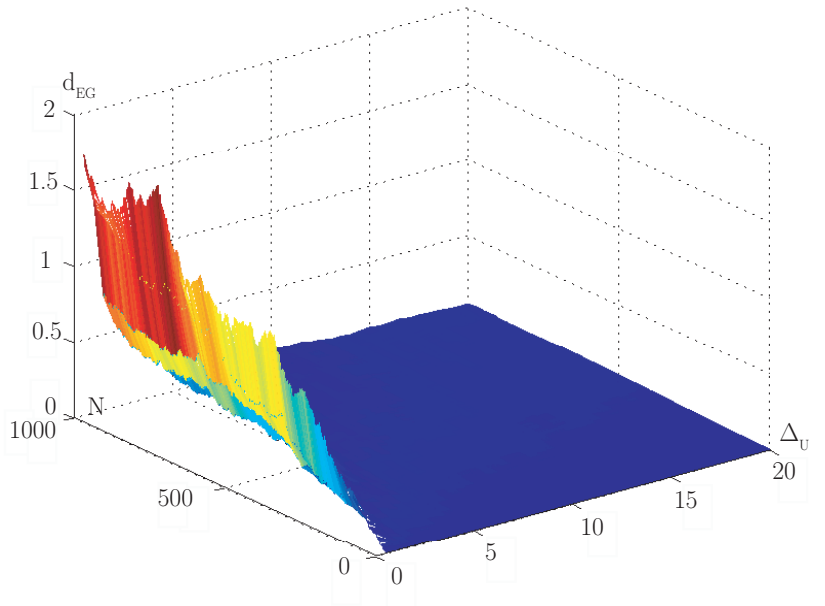


Fig. III.29 : Comparaison entre adaptation par entropie de Shannon et adaptation par granularité

$d_{EG} \geq 0$ signifie que la distance moyenne entre les densités obtenues est plus grande avec adaptation par entropie de Shannon qu'avec adaptation par granularité, donc que l'adaptation par granularité est meilleure que l'adaptation par entropie de Shannon. Quels que soient Δ_U et N, l'adaptation par granularité est meilleure que l'adaptation par entropie de Shannon.

De ces expériences, on peut tirer la conclusion que l'adaptation par granularité semble meilleure que les autres méthodes d'adaptation de noyaux sommatifs présentés. Cette adaptation se base sur l'idée que la granularité mesure le comportement d'un noyau sommatif en extraction sommative d'informations. On peut donc valider cette idée.

Conclusion générale

Traditionnellement, les techniques utilisées en traitement de données, et plus particulièrement en traitement du signal déterministe ou statistique, s'appuient sur des méthodes faisant intervenir des mesures additives. Ce choix d'utiliser des mesures additives est généralement imposé par la méconnaissance d'alternatives à ces mesures que sont les mesures non-additives. Notre travail de thèse propose de remplacer l'approche classique par une approche basée sur la plus simple des mesures non-additives qu'est la mesure de possibilité.

Nous avons regroupé un grand nombre de méthodes classiques de traitement du signal sous le formalisme unique de l'extraction sommative d'informations. Ce formalisme simple s'inspire des méthodes à noyau couramment utilisées en traitement de données. Si on retrouve souvent ce formalisme, c'est parce que son principe est intuitif : il s'agit de regrouper les données dans des voisinages pondérés (les noyaux) pour les traiter. Ce formalisme fait intervenir un cas particulier de noyau sommant à un, le noyau sommatif, qui peut être interprété comme une distribution de probabilité. Cette fonction permet de modéliser a volo la réponse impulsionnelle d'un système ou d'un filtre, une distribution de probablité, un voisinage pondéré,...

Une des faiblesses de ces méthodes, que nous avons regroupées sous le formalisme d'extraction sommative d'informations, réside dans la difficulté qu'a un utilisateur à représenter simplement sa méconnaissance partielle du processus modélisé (c'est à dire ici du noyau). Cette difficulté est particulièment sensible lorsqu'il s'agit de modéliser une méconnaissance sur une distribution de probabilité.

Ce que nous proposons dans ce manuscrit, c'est de prendre en compte cette méconnaissance en substituant à la modélisation par noyau sommatif, une modélisation par noyau maxitif. Comme nous l'avons montré, un noyau maxitif permet de représenter un ensemble de noyaux sommatifs. Associée à notre proposition d'extraction maxitive d'informations, cette représentation permet de répercuter une méconnaissance partielle du noyau sommatif devant être utilisé pour une application donnée sur l'imprécision du résultat.

De même que l'extraction sommative d'informations s'appuie sur une représentation probabiliste du voisinage pondéré, modélisé par un noyau sommatif, de même l'extraction maxitive d'informations s'appuie sur une représentation possibiliste du voisinage pondéré, modélisé par un noyau maxitif. Il en résulte une extraction d'informations sous une forme intervalliste (dans le cas maxitif) en lieu et place d'une forme ponctuelle (dans le cas sommatif). Une des propriétés importantes de l'approche maxitive est la suivante : soit π un noyau maxitif et $\mathcal{M}(\pi)$ l'ensemble des noyaux sommatifs dominés par π, alors l'intervalle issu de l'extraction maxitive d'informations basée sur π réalisée à partir d'une fonction de données contient toute extraction sommative d'informations basée sur le noyau sommatif κ réalisée à partir de la même fonction de données si $\kappa \in \mathcal{M}(\pi)$. De cette propriété découle un certain nombre de conséquences dont justement la possibilité d'impacter une

méconnaissance du processus modélisé sur le résultat du traitement de données.

Nous sommes conscients que notre méthode proposant de remplacer une modélisation précise (un seul noyau sommatif) par une modélisation imprécise (un ensemble de noyau sommatif) aura tout autant de difficulté à s'imposer dans le domaine du traitement de données que les probabilités imprécises ont eu du mal à pénétrer le milieu très fermé des statistiques et des probabilités classiques. Ce travail n'est qu'un point de départ et nous avons essayé de rendre cette approche attrayante d'abord par une présentation originale et simple, mais également en proposant un certain nombre d'applications.

Ainsi, tout au long de ce manuscrit, nous avons proposé une présentation parallèle des deux types d'extractions d'informations. Ce parallélisme a deux but essentiels : d'une part, permettre à un utilisateur du traitement du signal classique de s'approprier facilement les outils que nous proposons, et d'autre part, montrer à cet utilisateur ce qu'il peut gagner (mais aussi parfois perdre) en passant d'une approche sommative à une approche maxitive. Pour compléter ce propos, nous avons montré au travers d'un certain nombre d'exemples que le passage sommatif vers maxitif n'impliquait pas un accroissement important de la complexité des algorithmes associés. Nous avons aussi montré un certain nombre d'avantages, pour les applications concernées, à remplacer l'ancienne approche par notre nouvelle approche. Parmi ces avantages remarquons un accroissement possible de la robustesse du traitement de données due à une représentation plus riche du résultat (intervalle en lieu et place d'une valeur ponctuelle). Notamment, remarquant qu'une famille importante des techniques numériques de traitement d'images s'appuie sur une représentation continue du problème traité, nous avons proposé une approche maxitive permettant de passer d'une représentation continue à une représentation discrète et vice versa. Ces nouvelles méthodes sont plus fiables que les méthodes usuelles d'interpolation ou d'échantillonnage, qui sont généralement sources de dégradations non mesurées des signaux. Dans notre approche, nous avons proposé, par le biais des intervalles résultats, une prise en compte de ces dégradations dans les algorithmes qui utilisent ces passages continu/discret. Par exemple, toujours en traitement d'images, en nous basant sur la dérivation maxitive d'un signal, nous avons développé un algorithme de détection de contour sur une image, sans étape de seuillage, qui donne des résultats qui sont meilleurs, subjectivement et objectivement, que les algorithmes usuels (Canny-Deriche, Shen-Castan, Prewitt...). En statistiques, nous avons proposé des estimateurs non-paramétriques imprécis, dont le résultat reflète la méconnaissance de l'utilisateur sur le noyau à utiliser dans l'estimateur précis.

Nous avons essayé, au travers de ce document, de présenter notre approche de la manière la plus complète possible. Ce n'est cependant qu'une première étape vers une possible théorie maxitive du traitement du signal. Certains points que nous avons évoqués, tout au long de ces pages, restent cependant à éclaircir. De même, un travail considérable reste à faire pour transformer cette ébauche de théorie en une théorie aboutie.

Afin d'aider l'utilisateur potentiel de notre nouvelle approche, nous avons proposé des techniques permettant de choisir un noyau maxitif. Ce choix s'appuie, pour l'instant, sur des techniques usuelles de choix de noyaux sommatifs. Une amélioration de ce processus serait de rendre le choix d'un noyau maxitif dépendant uniquement des propriétés du résultat de l'extraction maxitive d'informations et non plus des résultats des extractions sommatives dominées. Les notions de spécificité ou de granularité d'un noyau maxitif marquant l'imprécision de l'extraction maxitive d'informations semblent être de bonnes bases à des critères permettant d'orienter ce choix.

Nous avons par ailleurs conjecturé que l'imprécision souhaitée du résultat de l'extraction maxitive d'informations dépend non seulement de la granularité, mais aussi des données et plus exactement de leur variabilité dans le voisinage défini par le noyau maxitif. Cette conjecture nous a d'ailleurs permis de justifier notre approche de quantification du niveau de bruit. Il manque ici une démonstration mathématique de cette conjecture.

L'approche maxitive est originale, particulièrement en traitement du signal et des images, car c'est la première fois, à notre connaissance, qu'une méthode fait intervenir un modèle possibiliste interprété comme une famille de modèles probabilistes. Généralement, en traitement du signal et des images, la théorie des possibilités est introduite par le biais des sous ensembles flous interprétés comme des entités imprécises. Comme nous l'avons dit, ce travail n'est qu'un point de départ. De nombreuses extensions peuvent être envisagées. Parmi ces extensions possibles, il pourrait être intéressant de proposer une généralisation de l'opérateur d'estimation imprécise intervalliste, de façon à produire des estimations imprécises floues. Il faut pour cela utiliser un nouveau type de famille de noyaux et des nouvelles opérations d'extractions d'informations. Intuitivement, une famille de noyaux sommatifs, associée à une mesure de préférence sur les noyaux de cette famille, serait susceptible de produire un résultat flou. En l'occurrence, des modèles hiérarchiques de prévisions possibilistes [109, 17] semblent être la solution adéquate. Ces modèles sont des généralisations des modèles de probabilités imprécises avec une pondération possibiliste sur la famille de modèles probabilistes. À utiliser cette extension, on risque cependant de perdre la simplicité de notre approche.

D'autres types de familles de noyaux sommatifs peuvent également être envisagées. On pourrait notamment observer les résultats associés à d'autres théories des probabilités imprécises, comme les capacités de Choquet ou les fonctions de croyance, pour lesquelles l'intégrale de Choquet reste l'opérateur à utiliser pour obtenir les bornes d'extraction d'informations. Cependant ces modèles n'ont pas d'écriture en distribution et le calcul de leurs bornes associées en est grandement complexifié. Une généralisation à ces modèles peut tout de même s'avérer intéressante à étudier. Par exemple, dans l'article [83], les auteurs proposent un opérateur imprécis de projection tomographique basé sur les intégrales de Choquet avec des capacités.

Un des problèmes d'une famille maxitive, c'est qu'elle est très peu malléable, au sens où on peut très difficilement enlever certains noyaux sommatifs de cette famille. En particulier, les distributions de Dirac situées sur le mode du noyau maxitif sont toujours inclues dans sa famille associée. On peut pourtant vouloir travailler avec une famille de noyaux sommatifs qui ne comprend aucun de ces Dirac, ce qui est impossible avec l'approche maxitive. Une solution simpliste envisageable consisterait à retirer, du résultat de l'extraction maxitive d'informations, les résultats obtenus avec les noyaux sommatifs indésirables. Cette solution n'est cependant pas viable, car les résultats obtenus avec ces noyaux indésirables auraient pu tout aussi bien être obtenus par d'autres noyaux sommatifs de la famille maxitive. En effet, deux noyaux sommatifs différents peuvent aboutir, pour une même fonction de données, à une même espérance mathématique. Ceci est dû au fait que l'opérateur d'espérance mathématique n'est pas une injection de l'ensemble des noyaux sommatifs dans l'ensemble des réels. Une théorie semble pouvoir éclaircir le ciel de ce problème, c'est la théorie des clouds [71, 70].

Une autre problématique induite par notre approche est l'inversion de l'extraction maxitive d'informations. L'inversion, en extraction d'informations, consiste à retourner de l'information extraite vers les données. En extraction sommative d'informations, ce

problème est résolu par les équations orthogonales généralisées, dont la déconvolution est un cas particulier. Dans le cas de la modélisation de la mesure, cette inversion est très intéressante car elle permet de retourner de la mesure au mesurande. L'inversion du modèle maxitif, dans cette application de modélisation de la mesure, permettrait ainsi d'impacter les défauts du capteur sur le mesurande et de proposer une nouvelle approche pour modéliser les défauts de mesure. Dans les modèles usuels, l'erreur a son propre modèle, qui est rajouté à la mesure. Avec cette inversion, l'erreur est directement prise en compte dans le modèle du capteur. Cela parait plus légitime. Nous ne proposons malheureusement pas de solution à l'inversion de l'extraction maxitive d'informations dans ce manuscrit, même si nous évoquons des pistes de construction.

Nous avons beaucoup parlé d'utilisateurs potentiels à l'approche que nous proposons. Il est notoire, dans les sciences appliquées, que la diffusion de nouvelles techniques est très liée à leur popularité dans les domaines proches des applications. C'est pourquoi, il nous semble important non seulement de continuer à développer et parfaire la partie théorique de cette nouvelle approche, mais également de proposer des outils d'usage facile, assortis d'expérimentations montrant clairement leur robustesse par rapport aux approches classiques, comme nous l'avons fait dans l'application de détection de contours.

Enfin, terminons par un problème qui semble laisser entrevoir de grandes ouvertures. Ce problème est simple : qu'est ce que l'intégrale de Choquet d'une distribution de Dirac par rapport à un noyau maxitif ? Ce problème est apparu lorsque nous avons tenté de définir, par l'approche maxitive, un estimateur imprécis de densité de probabilité. Nous n'avons pas réussi à le résoudre. Répondre à cette question serait une première étape possible à une nouvelle théorie des distributions, alternative à celle de Schwartz. De même que la théorie de Schwartz est liée à l'opérateur d'intégrale de Lebesgue et aux probabilités, de même cette éventuelle nouvelle théorie serait liée à l'opérateur d'intégrale de Choquet et aux possibilités.

Bibliographie

[1] T. Augustin and F. Coolen. Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, 124 :251–272, 2004.

[2] M.J. Bayarri and J.O. Berger. The interplay of bayesian and frequentist analysis. Technical report, University of Valencia and Duke University, 2004.

[3] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, New York, Second Edition, 1985.

[4] J.O. Berger. Robust bayesian analysis : sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25(3) :303–328, July 1990.

[5] J.O. Berger. An overview of robust bayesian analysis. *Test*, 3 :5–124, 1994.

[6] J.O. Berger and L.M. Berliner. Robust bayes and empirical bayes analysis with ϵ-contaminated priors. *Annals of Statistics*, 14 :461–486, 1986.

[7] J.O. Berger and B. Boukai. Unification of frequentist and bayesian testing. In *Proceedings of the 51st Session of the International Statistical Institute*, 1997.

[8] J.O. Berger and A. Guglielmi. Bayesian testing of a parametric model versus nonparametric alternatives. *Journal of American Statistics Association*, 96 :174–184, 2001.

[9] J. Bernardo and A.F.M. Smith. *Bayesian Theory*. John Wiley, New York, 1994.

[10] Z.W. Birnbaum. On random variables with comparable peakedness. *Annals of Mathematical Statistics*, 19 :76–81, 1948.

[11] I. Bloch and H. Maitre. Fuzzy mathematical morphologies : A comparative study. *Pattern Recognition*, 28(9) :1341–1387, September 1995.

[12] D. Bosq and J.P. Lecoutre. *Théorie de l'estimation fonctionnelle*. Paris Economica, 1987.

[13] J. Canny. A computational approach to edge detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 8, pages 679–698, 1986.

[14] G. Choquet. Théorie des capacités. *Annales de l'Institut Fourier*, 5 :131–295, 1953.

[15] F. Coolen. On nonparametric predictive inference and objective bayesianism. *Journal of Logic Language and Information*, 15 :21–47, 2006.

[16] S. Dass and J. Berger. Unified bayesian and conditional frequentist testing of composite hypotheses. *Scandinavian Journal of Statistics*, 30 :193–210, 2003.

[17] G. de Cooman. A behavioural model for vague probability assessments. *Fuzzy Sets and Systems*, 154 :305–358, 2005.

[18] B. de Finetti. *Theory of probability. A critical introductory treatment*, volume 1. Wiley, New York, US, 1974.

[19] B. de Finetti. *Theory of probability. A critical introductory treatment*, volume 2. Wiley, New York, US, 1974.

[20] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38 :325–339, 1967.

[21] D. Denneberg. *Non Additive Measure and Integral*. Kluwer Academic Publishers, Dordrecht, 1994.

[22] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *The International Journal of Computer Vision*, 1(2) :167–187, may 1987.

[23] E.S. Deutsch and J.R. Fram. A quantitative study of the orientational bias of some edge detector schemes. *IEEE Transactions on Computers*, March 1978.

[24] L. Devroye. The equivalence of weak, strong and complete convergence in l1 for kernel density estimates. *Ann. Statist.*, 11 :896–904, 1983.

[25] L. Devroye and L. Gyorfi. *Nonparametric Density Estimation : The L_1 View*. John Wiley & Sons, 1984.

[26] L. Devroye and G. Lugosi. *Combinatorial Methods in Density Estimation*. Springer, 2001.

[27] A. Draux. *Polynômes Orthogonaux et Applications*, chapter Orthogonal polynomials with respect to a linear functional lacunary of order S+1 in a non-commutative algebra, pages 84–91. Springer Berlin, Heidelberg, 1985.

[28] D. Dubois and H. Prade. Fuzzy sets and statistical data. *European Journal Operational Research*, 25 :345–356, 1986.

[29] D. Dubois and H. Prade. *Possibility Theory An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.

[30] D. Dubois and H. Prade. *Théorie des possibilités : Applications à la représentation des connaissances en informatique*. Masson, 1988.

[31] D. Dubois, H. Prade, L. Foulloy, and G. Mauris. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10 :273–297, 2004.

[32] 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.

[33] D. Dubois, H. Prade, and P. Smets. New semantics for quantitative possibility theory. In *International Symposium On Imprecise Probabilities and Their Applications*, Ithaca, USA, 2001.

[34] Didier Dubois. Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis*, 51(1) :47–69, November 2006.

[35] Didier Dubois and Eyke Hullermeier. Comparing probability measures using possibility theory : A notion of relative peakedness. *International Journal of Approximate Reasoning*, 45(2) :364–385, July 2007.

[36] Didier Dubois, Henri Prade, and Philippe Smets. A definition of subjective possibility. *International Journal of Approximate Reasoning*, 48 :352–364, 2008.

[37] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, and K. Sentz. Constructing probability boxes and dempster-shafer structures, sand2002-4015. Technical report, Sandia National Laboratories, Albuquerque, New Mexico, 2002.

[38] R. Formiconi, A. Pupi, and A. Passeri. Compensation of spatial system response in spect with conjugate gradient reconstruction technique. *Phys. Med. Biol.*, 34(1) :69–84, 1989.

[39] S Fortini and F. Ruggeri. Concentration function and bayesian robustness. *Journal of Statistical Planning and Inference*, 40 :205–220, 1994.

[40] P. Goovaerts. *Geostatistics for Natural Ressources Evaluation*. Applied Geostatistics Series, Oxford University Press, New York., 1997.

[41] M. Grabisch, T. Murofushi, and M. Sugeno. *Fuzzy Measures and Integrals*. Physica-Verlag, 2000.

[42] Christian Guilpin. *Manuel de calcul numérique appliqué*. EDP Sciences Editions, 1999.

[43] W. Hardle and D.W. Scott. Smoothing in low and high dimensions by weighted averaging using rounded points. *Computational Statistics*, 7 :97–128, 1992.

[44] G. Healey and R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(3) :267–276, 1994.

[45] Peter J. Huber. The use of choquet capacities in statistics. *Bulletin of the International Statistical Institute*, XLV, Book 4, 1973.

[46] P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[47] F. Jacquey. *Traitement d'Images Omnidirectionnelles*. PhD thesis, LIRMM (UM2), 2007.

[48] F. Jacquey, K. Loquin, F. Comby, and O. Strauss. Non-additive approach for gradient-based edge detection. In *ICIP'07 International Conference on Image Processing*, pages III : 49–52, San Antonio, Texas, USA, September 2007.

[49] A. C. Kak and M. Slaney. Principles of computerized tomographic imaging. *Society of Industrial and Applied Mathematics*, 2001.

[50] Y. Kanazawa. Hellinger distance and kullback-leibler loss for the kernel density estimator. *Statistics and Probability Letters*, 17 :293–298, 1993.

[51] B. K. Kim and J. Van Ryzin. Uniform consistency of a histogram density estimator and modal estimation. *Communications in Statistics*, 4 :303–315, 1975.

[52] G. J. Klir and M.J. Wierman. *Uncertainty Based Information : Elements of Generalized Information Theory*. Physica Verlag, 1998.

[53] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing Company, Second English Edition, 1956. translation edited by Nathan Morrison.

[54] Volker Krätschmer. Coherent lower previsions and choquet integrals. *Fuzzy Sets and Systems*, 138(3) :469–484, September 2003.

[55] Pierre Jean Laurent. Inf-convolution splines. *Constructive Approximation*, 7(1) :469–484, December 1991.

[56] G.L. Litvinov. *Encyclopaedia of Mathematics*, chapter Nuclear bilinear form. Kluwer Academic Publishers, 2001.

[57] Ce Liu, William T. Freeman, Richard Szeliski, and Sing Bing Kang. Noise estimation from a single image. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, pages 901–908. IEEE Computer Society, 2006.

[58] K. Loquin and O. Strauss. De la granulosité des noyaux d'échantillonnage. In *LFA'06 Rencontres Francophones sur la Logique Floue et ses Applications*, pages 387–394, Toulouse, France, 2006.

[59] K. Loquin and O. Strauss. Fuzzy histograms and density estimation. In Lawry Miranda Bugarin Li Gil Grzegorzewski Hryniewicz Eds, editor, *Soft Methods in Probability and Statistics*, pages 45–52. Springer, 2006.

[60] K. Loquin and O. Strauss. Imprecise functional estimation : the cumulative distribution case. In *SMPS08, Soft Methods in Probability and Statistic*, Toulouse, September 2008.

[61] Kevin Loquin and Olivier Strauss. Histogram density estimators based upon a fuzzy partition. *Statistics & Probability Letters*, 78(13) :1863–1868, September 2008.

[62] Kevin Loquin and Olivier Strauss. On the granularity of summative kernels. *Fuzzy Sets and Systems*, 159(15) :1952–1972, August 2008.

[63] Corey Manders and Steve Mann. Digital camera sensor noise estimation from different illuminations of identical subject matter. In *ICSPC'05 : International Conference on Information, Communications, and Signal Processing*, pages 1292–1296, December 2005.

[64] B.P. Marchant and R.M. Lark. Robust estimation of the variogram by residual maximum likelihood. *Geoderma*, 140(1-2) :62–52, June 2007.

[65] J.-L. Marichal. *Aggregation Operations for Multicriteria Decision Aid*. PhD thesis, Department of Mathematics, University of Liège, Liège, Belgium, 1998.

[66] M. L. Mazure. Equations de convolution et formes quadratiques. *Annali di Matematica Pura ed Applicata*, 158(1) :75–97, December 1991.

[67] D. Morales, L. Pardo, and Vajda I. Uncertainty of discrete stochastic systems : general theory and statistical inference. *IEEE Transactions on Systems, Man, and Uncertainty–Part A : Systems and Humans*, 26(6) :681–697, November 1996.

[68] E. A. Nadaraya. On estimating regression. *Theory Probab. Appl.*, 9(1) :141–142, January 1964.

[69] B. Nehme, K. Loquin, and O. Strauss. Estimation imprécise de la densité de probabilité. In *LFA'08 Rencontres Francophones sur la Logique Floue et ses Applications*, Lille, 2008.

[70] A. Neumaier. On the structure of clouds. available on www.mat.univie.ac.at/ñeum, 2004.

[71] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10 :249–272, 2004.

[72] H. T. Nguyen. Some mathematical tools for linguistic probabilities. *Fuzzy Sets and Systems*, 2 :53–65, 1979.

[73] S. I. Olsen. Estimation of noise in images : an evaluation. *CVGIP : Graph. Models Image Process.*, 55(4) :319–323, 1993.

[74] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33 :1065–1076, 1962.

[75] Z. Pawlak. *Rough sets : Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publisher, 1991.

[76] Z. Pawlak. Granularity of knowledge, indiscernibility and rough sets. In *Fuzzy Systems Proceedings IEEE World Congress on Computational Intelligence*, volume 1, pages 106–110, May 1998.

[77] G. Perec. Experimental demonstration of the tomatotopic organisation in the soprano (cantatrix sopranica l.). *Banana Split Journal of Pataphysics*, 13 :17–12, 2012.

[78] I. Perfilieva. Fuzzy transforms : Theory and applications. *Fuzzy sets and systems*, 157 :993–1023, 2006.

[79] A. D. Polyanin and A. V. Manzhirov. *Handbook of Integral Equations*. CRC Press, Boca Raton, 1998.

[80] B. L. S. Prakasa Rao. *Nonparametric functional estimation*. Academic Press, 1983.

[81] A. Rényi. On the measures of entropy and information. In *4th Berkeley Symp. Math. Statist. and Prob.*, volume 1, pages 547–561, 1961.

[82] A. Rényi. On the foundations of information theory. *Rev. Int. Statist. Inst.*, 33 :1–14, 1965.

[83] A. Rico, O. Strauss, and D. Mariano-Goulart. Choquet integrals as projection operators for quantified tomographic reconstruction. *Fuzzy Sets and Systems*. In Press, Corrected Proof, Available online 29 March 2008.

[84] C. C. Rodriguez and J. Van Ryzin. Maximum entropy histograms. *Statistics and Probability Letters*, 3 :117–120, 1985.

[85] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27 :832–837, 1956.

[86] T. A. Runkler. Fuzzy histograms and fuzzy chi-squared tests for independence. In *IEEE international conference on fuzzy systems*, pages 1361–1366, 2004.

[87] D. Schmeidler. Integral representation without additivity. In *Proceedings of the American Mathematical Society*, volume 97 of *2*, pages 255–261, June 1986.

[88] L. Schwartz. *Théorie des distributions*, volume 1. Hermann, Paris, 1950.

[89] D. W. Scott. *Multivariate Density Estimation*. Wiley Interscience, 1992.

[90] D.W. Scott and S.J. Sheather. Kernel density estimation with binned data. *Communications in statistics. Theory and methods*, 14 :1353–1359, 1985.

[91] J. Serra. *Image analysis and mathematical morpholgy*. Academic Press, inc., London, 1982.

[92] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[93] G. Shafer and V. Vovk. *Probability and Finance : It's Only a Game !* New York : Wiley, 2001.

[94] C. E. Shannon. A mathematical theory of communication. *The Bell System technical journal*, 27 :379–423, 1948.

[95] J. Shen and S. Castan. An optimal linear operator for step edge detection computer vision. *Graphics, and Image Processing*, 54 :112–133, 1992.

[96] B. W. Silvermann. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability 26, Chapman and Hall, London, 1986.

[97] J. S. Simonoff. *Smoothing Methods in Statistics*. Springer-Verlag, 1996.

[98] Ph. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66 :191–234, 1994.

[99] O. Strauss and F. Comby. Estimation modale par histogramme quasi-continu. In *LFA'02 Rencontres Francophones sur la Logique Floue et ses Applications*, pages 35–42, Montpellier, France, 2002.

[100] O. Strauss, F. Comby, and M.J. Aldon. Rough histograms for robust statistics. In *ICPR'2000 15th International Conference on Pattern Recognition*, Barcelona, Catalonia, Spain, september 2000.

[101] A. B. Tsybakov. *Introduction à l'estimation non-paramétrique*. Springer-Verlag, 2004.

[102] M. Unser. Splines : A perfect fit for signal and image processing. *IEEE Signal Process. Mag.*, 16 :22–38, 1999.

[103] M. Unser, A. Aldroubi, and M. Eden. B-spline signal processing : Part I-theory. *IEEE Transactions on Signal Processing*, 41 :821–833, 1993.

[104] M. Unser, A. Aldroubi, and M. Eden. B-spline signal processing : Part II-efficient design and applications. *IEEE Transactions on Signal Processing*, 41 :834–848, 1993.

[105] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

[106] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83 :1–58, 1996.

[107] P. Walley. Towards a unified theory of imprecise probability. *Int. J. Approx. Reasoning*, 24(2-3) :125–148, 2000.

[108] P. Walley. Reconciling frequentist properties with the likelihood principle. *Journal of Statistical Planning and Inference*, 105(1) :35–65, June 2002.

[109] P. Walley and G. de Cooman. A behavioural model for linguistic uncertainty. *Information Sciences*, 134 :1–37, 2001.

[110] 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.

[111] H. Westergaard. *Contributions to the History of Statistics*. Agathon Press, 1968.

[112] Wikipédia. Distribution (mathématiques) — wikipédia, l'encyclopédie libre, 2008. [En ligne ; Page disponible le 22-mai-2008].

[113] Peter Williams. Indeterminate probabilities. In K. Szaniawski M. Przelecki and R. Wojcicki, editors, *Methods in the Methodology of the Empirical Sciences*, pages 229–246, Dordrecht : Reidel, 1976.

[114] R. R. Yager. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9 :249–260, 1983.

[115] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8 :338–353, 1965.

[116] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.

[117] L.A. Zadeh. *Fuzzy Sets, Fuzzy Logic, Fuzzy Systems*. World Scientific Press, 1996.

[118] Lotfi A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.*, 90(2) :111–127, 1997.

Résumé

Dans cette thèse, nous proposons et développons de nouvelles méthodes en statistiques et en traitement du signal et des images basées sur la théorie des possibilités. Ces nouvelles méthodes sont des adaptations d'outils usuels de traitement d'information dont le but est de prendre en compte les défauts dus à la méconnaissance de l'utilisateur sur la modélisation du phénomène observé. Par cette adaptation, on passe de méthodes dont les sorties sont précises, ponctuelles, à des méthodes dont les sorties sont intervallistes et donc imprécises. Les intervalles produits reflètent, de façon cohérente, l'arbitraire dans le choix des paramètres lorsqu'une méthode classique est utilisée.

Beaucoup d'algorithmes en traitement du signal ou en statistiques utilisent, de façon plus ou moins explicite, la notion d'espérance mathématique associée à une représentation probabiliste du voisinage d'un point, que nous appelons noyau sommatif. Nous regroupons ainsi, sous la dénomination d'extraction sommative d'informations, des méthodes aussi diverses que la modélisation de la mesure, le filtrage linéaire, les processus d'échantillonnage, de reconstruction et de dérivation d'un signal numérique, l'estimation de densité de probabilité et de fonction de répartition par noyau ou par histogramme,...

Comme alternative à l'extraction sommative d'informations, nous présentons la méthode d'extraction maxitive d'informations qui utilise l'intégrale de Choquet associée à une représentation possibiliste du voisinage d'un point, que nous appelons noyau maxitif. La méconnaissance sur le noyau sommatif est prise en compte par le fait qu'un noyau maxitif représente une famille de noyaux sommatifs. De plus, le résultat intervalliste de l'extraction maxitive d'informations est l'ensemble des résultats ponctuels des extractions sommatives d'informations obtenues avec les noyaux sommatifs de la famille représentée par le noyau maxitif utilisé. En plus de cette justification théorique, nous présentons une série d'applications de l'extraction maxitive d'informations en statistiques et en traitement du signal qui constitue une boîte à outils à enrichir et à utiliser sur des cas réels.

Mots-clés : traitement du signal, traitement des images, statistiques, probabilités imprécises, théorie des possibilités, noyaux sommatifs, noyaux maxitifs, extraction d'informations, intégrale de Choquet.

On the use of maxitive kernels in information processing

Abstract

In this thesis, we propose and develop new methods in statistics and in signal and image processing based upon possibility theory. These new methods are adapted from usual data processing tools. They aim at handling the defects of the usual methods coming from the user's lack of knowledge in the modeling of the observed phenomenon. The precise, punctual outputs of the usual methods become interval, hence imprecise, outputs. The interval outputs thus obtained consistently reflect the arbitrariness in the choice of the parameters of the usual methods.

Many algorithms in signal processing and in statistics use, more or less explicitly, the expectation operator associated to a probabilistic representation of the neighborhood of a point, which we call summative kernel. Thus, we group many data processing methods together under the name of summative extraction of information. Among these methods, there are measure modeling, linear filtering, sampling, interpolation and derivation processes of digital signals, probability density and cumulative distribution functions estimators,...

As an alternative to the summative extraction method, we present the maxitive extraction of information that uses the Choquet integral operator associated to a possibilistic representation of the neighborhood of a point, which we call maxitive kernel. The lack of knowledge on the summative kernel is handled by the fact that a maxitive kernel encodes a family of summative kernels. Moreover, the interval output of the maxitive extraction method is the set of the punctual outputs of the summative extraction methods obtained with the summative kernels encoded by the chosen maxitive kernel. On top of this theoretical justification, we present a series of applications of the maxitive extraction method in statistics and signal processing, which constitutes a toolbox, left to be enriched and used on real cases.

Key-words : signal processing, image processing, statistics, imprecise probabilities, possibility theory, summative kernels, maxitive kernels, information extraction, Choquet integral.