On the granularity of summative kernels

Kevin Loquin, Olivier Strauss

On the granulosity of summative kernels

Kevin Loquin [*], Olivier Strauss

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

Abstract

In this paper, we propose *granulosity* as a new index to characterize the non-specificity of a summative kernel. This index is intended to reflect the behavior of a kernel in the usual signal processing applications. We show, in different experiments, that two kernels having the same granulosity have very similar behavior. This granulosity-based adaptation is compared to other adaptation methods. These experiments highlight the ability of the granulosity index to measure the spreading and collecting properties of a summative kernel.

Key words: summative, maxitive kernels, granulosity, possibility, probability theory, signal processing, adaptation

1 Introduction

A wide range of digital analysis and signal processing [16] procedures inherently rely on methods for reconstructing a continuous underlying signal from a set of sampled corrupted values. These values are usually uniformly sampled, since the measures come from systematic observations. Kernels are essential tools in this context, since they are used in reconstruction, impulse response modelling, resampling, interpolation, linear or non-linear transformations, stochastic or band-pass filtering, etc. In digital signal processing, kernels are mainly used to derive discrete algorithms from a continuous representation. Within most applications, a kernel can be seen as a weighted neighborhood ensuring a smooth interplay between continuous and discrete domains. They can be visualized as bumps that can be shifted to any location

---

[*] Corresponding author. Tel: +33 467 41 85 65; fax: +33 467 41 85 00
  Email addresses: `kevin.loquin@lirmm.fr` (Kevin Loquin),
`olivier.strauss@lirmm.fr` (Olivier Strauss).

of the signal domain, so as to absorb or spread the information contained in the signal. They are often bounded, monomodal and symmetric. In this paper, the kernels will be defined on a domain Ω, subset of $\mathbb{R}^p$, for $p \in \mathbb{N}$.

Digital signal derivation [15, 3, 5] is a typical example of such an application. The classical finite differences method usually fails to perform the estimation of the derivative of the signal, especially with a noisy signal. The kernel-based method consists of computing the sampled derivative of an estimation of the continuous signal. This estimation is obtained by convolving the original discrete signal with a continuous kernel, chosen to lower the impact of both acquisition noise and quantization effect. The implementation of such a method simply consists of convolving the original sampled signal with the derivative of the chosen kernel, which is also a kernel.

Most of the kernels used in signal processing are summative kernels, or linear combinations of summative kernels. A summative kernel is a positive function, the integral of which equals one. For instance splines [32] are summative kernels, the derivatives of which are linear combinations of splines.

One problem of practical importance is the characterization of the behavior of a summative kernel in a particular application. Our aim, in this paper, is to define an index reflecting the ability of a summative kernel to absorb or spread the information contained in the signal. This is particularly sensible when the summative kernel is the impulse response of a sensor. In that case, this index should capture the non-resolution power of the sensor, i.e. its power of information collecting. Note that since the neighborhood, represented by the kernel, can be shifted to different locations of the signal, its non-resolution power should not depend on its location. When considering a simple crisp neighborhood, its non-resolution power can be naturally quantified by its Lebesgue measure, which is a continuous generalization of the concept of cardinality of a discrete set. This quantification of non-resolution power of a crisp neighborhood can easily be generalized to fuzzy neighborhoods by extending the concept of cardinality to the continuous domain. Cardinality is an index of the non-specificity of a fuzzy neighborhood. Specificity, as stated in [8, 35], can be considered as a distance measure between the fuzzy neighborhood considered and a fuzzy set that contains only one element. It can be reformulated in the following manner: the specificity of a fuzzy neighborhood is its ability to be concentrated on a set of minimal length (in the sense of the Lebesgue measure of this set), i.e. its ability to be influent on specific parts of the domain of the neighborhood. However, no such index has been defined for a summative kernel-based weighted neighborhood.

Since a summative kernel, as defined above, can also be seen as a probability distribution, it seems natural to consider existing information measures of probability distribution, regarding their relevance as indices of non-resolution

power. The notion of informativity, for a probability distribution, corresponds to the information needed to construct it. In the literature, an information measure of a probability distribution is supposed to quantify its informativity. This quantification is however sometimes doubtful, since what is captured by an information measure of a probability distribution is, rather, the geometrical properties of the distribution, such as its flatness, its peakedness, or its average difference from the mean. To take a simple example: uniform distribution, whether obtained under full knowledge (i.e. where it is known that every alternative of Ω has the same probability of occuring) or under total indeterminacy and by the use of the principle of insufficient reason (i.e. where nothing is known about the probabilities of the alternatives of Ω), will have the same shape and the same information measure, whereas the informativity is different.

In this paper, we propose a non-resolution power index for a summative kernel-based weighted neighborhood under the name of granulosity, by reference to the works of Pawlak [21], who defined the granularity of a rough set as its power of resolution, and to the more recent works of Bodjanova [2], who defined a concept of granulation for a fuzzy set.

This paper is organized as follows. After this introduction, section 2 presents the concepts of maxitive and summative kernels. Each kernel is associated with a model of uncertainty: possibility and probability measures respectively. Considering each model, we present different ways of defining the non-specificity of a weighted neighborhood. Section 3 is dedicated to the construction of granulosity as an index of non-specificity. Section 4 presents a granulosity-based method for adapting kernels. This method is compared to other adaptation methods. In section 5 we expose illustrative experiments highlighting the ability of granulosity to measure the spreading and collecting properties of a summative kernel.

2   Kernels, uncertainty and information

2.1   Maxitive kernels, possibility distributions and normalized fuzzy subsets

What will be called a maxitive kernel in this paper is a possibility distribution or membership function of a normalized fuzzy subset [36]. It can be considered as a weighted neighborhood, dispersing or collecting information around a given location, called the mode. The mode can be a singleton or a set. Furthermore, the extent of this neighborhood is delimited by its support.

Definition 1 *A maxitive kernel is a $[0,1]$-valued function π, defined on a*

domain Ω, verifying the maxitivity property

$$\sup_{\omega \in \Omega} \pi(\omega) = 1. \tag{1}$$

Its mode is given by $Mode(\pi) = \{\omega \in \Omega / \pi(\omega) = 1\}$ and its support is $Supp(\pi) = \{\omega \in \Omega / \pi(\omega) > 0\}$.

Note that any given monomodal maxitive kernel π, defined on Ω, can be the basis for a family of maxitive kernels tuned by a location-scale parameter $\theta = (m, \Delta)$, with m a translation factor and Δ its bandwidth. Any element of this family is obtained, for $m \in \Omega$ and $\Delta > 0$, by

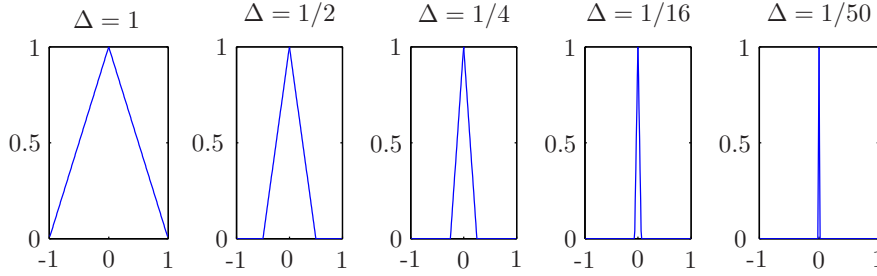$$\pi_\Delta^m(\omega) = \pi(\frac{\omega - m}{\Delta}), \ \forall \omega \in \Omega + \{m\}. \tag{2}$$

Figure 1. Triangular family of maxitive kernels

Figure 1 consists of 5 maxitive kernels π_Δ^0 of the family that has for its basic maxitive kernel the Triangular kernel $\pi(\omega) = (1 - |\omega|)\mathbb{1}_{[-1,1]}(\omega)$.

Note that figure 1 is in agreement with theorem 2, which tells us that the Kronecker function δ^m is the limit of a sequence of maxitive distributions whose graphs become thin and peaked around $\omega = m$. The Kronecker function is defined by $\delta^m(\omega) = 1$ for $\omega = m$ and 0 otherwise.

Theorem 2 *For a fixed $m \in \Omega$, let π_Δ^m be monomodal maxitive kernels on Ω, with mode m, such that for any $a > 0$, $\sup_{B_a} \pi_\Delta^m(\omega) \to 0$ as $\Delta \to 0$, with $B_a = \{\omega \in \Omega / \forall i, |\omega_i - m_i| \geq a\}$, then $\pi_\Delta^m \longrightarrow \delta^m$, as $\Delta \longrightarrow 0$, in the sens that $\lim_{\Delta \to 0} \sup_\Omega(\pi_\Delta^m(\omega)\varphi(\omega)) = \sup_\Omega(\delta^m(\omega)\varphi(\omega)) = \varphi(m)$, whatever φ a posibility distribution.*

Proof 3 *We note $A = \lim_{\Delta \to 0} \sup_\Omega(\pi_\Delta^m(\omega)\varphi(\omega))$. First, we have that $\Omega = \cup_{a>0}B_a \cup \{m\}$, then $A = \lim_{\Delta \to 0} \sup_{\cup_{a>0}B_a \cup \{m\}}(\pi_\Delta^m(\omega)\varphi(\omega))$. Since for any $a > 0$, $\lim_{\Delta \to 0} \sup_{B_a} \pi_\Delta^m(\omega) = 0$, the supremum of the last expression is obtained for $\omega = m$, therefore, $A = \lim_{\Delta \to 0} \pi_\Delta^m(m)\varphi(m) = \varphi(m)$, because whatever m and Δ, $\pi_\Delta^m(m) = 1$. $\square$*

4

A possibility distribution π has a relevant meaning in the scope of uncertainty theories. π induces a possibility measure, noted Π, computed in this way:

$$\forall A \subseteq \Omega, \ \Pi(A) = \sup_{\omega \in A} \pi(\omega), \tag{3}$$

which verifies the maxitivity axiom for possibility distributions defined on an infinite domain [19]:

$$\forall (A_n)_{n>0} \subseteq \Omega, \ \Pi(\bigcup_{n>0} A_n) = \max_{n>0}(\Pi(A_n)). \tag{4}$$

The value $\Pi(A)$ can be interpreted as a degree of possibility for a realization of the underlying uncertain phenomenon to fall in A. As stated in [34, 4, 9], a possibility measure is one particular case of upper probability, a notion introduced by Peter Walley [33]. Note that we can obtain the possibility distribution π from its measure Π on singletons by $\pi(\omega) = \Pi(\{\omega\})$, $\forall \omega \in \Omega$.

2.2 Construction of possibility distributions from orderings

The notion of degree is parallel to the notion of ordering. Indeed, every element, whose degree of appropriateness to a concept is given, can be ordered in a numerical scale. Conversely, every concept, according to which elements may be ordered, induces degrees of appropriateness of these elements to this concept. To take the example of mammals, an elephant can be thought of as being more a mammal than a duck-billed platypus, because the latter lays eggs, which is not a feature of mammals. When the underlying concept depends on numerical values, as for instance the concept of "tall" depends on height, it is more sensible and natural to build a numerical scale of degrees from the ordering induced by the heights.

Therefore, any ordering on alternatives (the ω of Ω) numerically scaled by non-normalized degrees $d = (d(\omega))_{\omega \in \Omega}$ can induce a fuzzy subset, whose membership function is a set of degrees on the alternatives. To comply with possibility theory, these non-normalized degrees d induce a set function $D(A) = \sup_{\omega \in A} d(\omega)$, $\forall A \subseteq \Omega$, which is not normalized either. For consistency of this ordering approach with the possibilistic interpretation of D, it is natural to assume that the degree of possibility that any realization of the uncertain phenomenon will fall in Ω is 1, i.e. that the post-normalized possibility measure, noted Π, fulfils the maxitivity property (1), which is equivalent to $\Pi(\Omega) = 1$. A simple procedure for achieving this normalization consists of dividing all the degrees d by $D(\Omega) = \sup_{\omega \in \Omega} d(\omega)$. The post-normalized degrees of possibility are $\pi(\omega) = \frac{d(\omega)}{D(\Omega)}$, $\forall \omega \in \Omega$. This normalization is also consistent when interpreting π as a maxitive kernel, i.e. the weighted neighborhood of a location.

5

Indeed, the mode is a set of locations fully in accordance with the concept that the neighborhood represents.

In a more formal way, for a set of alternatives $\omega \in \Omega$, a preference ordering $\succeq_\pi$ is equivalent to a possibility distribution π by:

$$\omega_1 \succeq_\pi \omega_2 \iff \pi(\omega_1) \geq \pi(\omega_2). \tag{5}$$

Preferring one alternative to another is equivalent to saying that one alternative is more possible than another.

2.3 Possibility and non-specificity

First, the concept of specificity already exposed in the literature [35, 8] of fuzzy sets and possibility theories should be remembered. Specificity is the ability for a fuzzy neighborhood to be concentrated on a set of minimal length (in the sense of the Lebesgue measure of this set). Measures of specificity effectively capture the idea of how close a fuzzy set is to a singleton. The Kronecker function at any given location ω of Ω is the most specific possibility distribution, whereas the vacuous possibility distribution on Ω, which equals 1 on Ω, is the least specific distribution. The first part of this remark, associated with theorem 2, leads to noting that for a family of maxitive kernels $(\pi_\Delta^m)_{\Delta > 0}$, given for any $m \in \Omega$, the specificity of π_Δ^m increases when $\Delta \longrightarrow 0$, since this family tends to the Kronecker function.

The continuous generalization of the cardinality of a fuzzy subset F, whose membership function is $\mu_F(\omega) = \pi(\omega)$, $\forall \omega$, is defined by expression (6).

$$card(\pi) = \int_\Omega \pi(\omega)d\omega. \tag{6}$$

This is a natural measure of its non-specificity.

Note that when π is interpreted as a maxitive kernel, its cardinality is a natural measure of its non-resolution power.

2.4 Summative kernels and probability distributions

What is called a summative kernel, in this paper, is a probability distribution. As for a maxitive kernel, it can be considered as a weighted neighborhood, dispersing or collecting information around a given location, called the mode. The mode can be a singleton or a set. Furthermore, the extent of this neighborhood is delimited by its support.

Definition 4 *Summative kernels are $\mathbb{R}^+$-valued functions κ defined on a domain Ω, verifying the summativity property*

$$\int_\Omega \kappa(x)dx = 1. \tag{7}$$

The mode of a summative kernel κ, noted $Mode(\kappa)$, are the values ω at which κ attains its maximum value. Its support is given by $Supp(\kappa) = \{\omega \in \Omega/\kappa(\omega) > 0\}$.

Note that any given monomodal summative kernel κ, can be the basis for a family of summative kernels tuned by a location-scale parameter $\theta = (m, \Delta)$, with m a translation factor and $\Delta > 0$ its bandwidth. Any element of this family is obtained by

$$\kappa_\Delta^m(\omega) = \frac{1}{\Delta}\kappa(\frac{\omega - m}{\Delta}), \ \forall \omega \in \Omega. \tag{8}$$

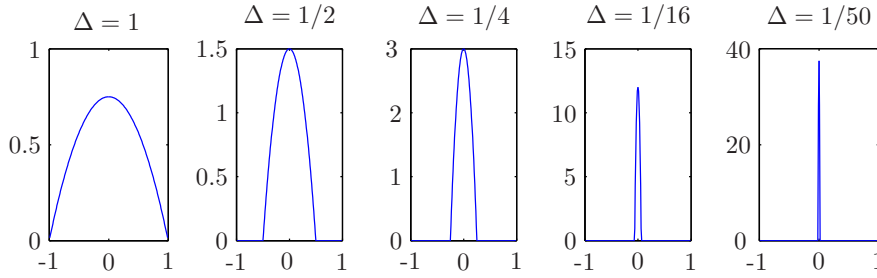
Figure 2. Epanechnikov family of summative kernels

Figure 2 consists of 5 summative kernels κ_Δ^0 of the family that has for its basic kernel the Epanechnikov kernel $\kappa(\omega) = \frac{3}{4}(1 - \omega^2)\mathbb{1}_{[-1,1]}(\omega)$.

Note that figure 2 is in agreement with theorem 5, seen in [26], which tells us that the Dirac delta function δ^m is the limit, in the sense of distributions, of a sequence of functions whose graphs become thin and tall, peaking around $\omega = m$. The Dirac delta function, often referred to as the unit impulse function, has the value of $+\infty$ for $\omega = m$ and the value zero elsewhere. For this theorem, we take $\Omega = \mathbb{R}$.

Theorem 5 *For a fixed $m \in \Omega$, let κ_Δ^m be monomodal summative kernels on Ω, with mode m, such that for any $a > 0$, $\int_{B_a} \kappa_\Delta^m(\omega)d\omega \to 0$ as $\Delta \to 0$, with $B_a = \{\omega \in \Omega/\forall i, |\omega_i - m_i| \geq a\}$, then $\kappa_\Delta^m \longrightarrow \delta^m$, as $\Delta \to 0$ in the sense of the distributions, i.e. $\lim_{\Delta \to 0} \int_\Omega \kappa_\Delta^m(\omega)\varphi(\omega)d\omega = \int_\Omega \delta^m(\omega)\varphi(\omega)d\omega = \varphi(m)$, whatever φ a test function.*

As a probability distribution, κ has a relevant meaning in the scope of uncertainty theories. It induces a probability measure, noted P, computed in this

way

$$\forall A \subset \Omega, \ P(A) = \int_A \kappa(\omega)d\omega, \tag{9}$$

which verifies the Kolmogorov additivity axiom, which says that for any count-able sequence of pairwise disjoint events $(A_n)_{n>0} \subseteq \Omega$,

$$P(\bigcup_{n>0} A_n) = \sum_{n>0} P(A_n). \tag{10}$$

The value $P(A)$ can be interpreted as a degree of probability for a realization of the underlying uncertain phenomenon to fall in A.

2.5 Construction of probability distributions from orderings

In the same way that a possibility distribution π can be obtained from a pref-erence ordering $\succeq_\pi$ on the alternatives, a probability distribution κ can be obtained from a preference ordering $\succeq_\kappa$ on the alternatives. In this sense, pre-ferring one alternative to another is equivalent to saying that one alternative is more probable than another.

To comply with probability theory, the non-normalized degrees, noted $d = (d(\omega))_{\omega \in \Omega}$, as in part 2.2, induce, under the condition that $d \in \mathrm{L}_1$, a set func-tion $D(A) = \int_A d(\omega)d\omega, \ \forall A \subseteq \Omega$, which is not normalized either. The nor-malization procedure is different from the possibilistic, because of the sum-mativity property (7). For consistency of this preference ordering approach with the probabilistic interpretation of D, it is natural to assume that the degree of probability that any realization of the uncertain phenomenon will fall in Ω is 1, i.e. that the post normalized probability measure, noted P, ful-fils the summativity property (7), which is equivalent to $P(\Omega) = 1$. A simple procedure for achieving this normalization consists of dividing all the degrees d by $D(\Omega) = \int_\Omega d(\omega)d\omega$. The post-normalization degrees of probability are $\kappa(\omega) = \frac{d(\omega)}{D(\Omega)}, \ \forall \omega \in \Omega$.

In a more formal way, for a set of alternatives $\omega \in \Omega$, a preference ordering $\succeq_\kappa$ is equivalent to a probability distribution κ by:

$$\omega_1 \succeq_\kappa \omega_2 \iff \kappa(\omega_1) \geq \kappa(\omega_2). \tag{11}$$

The preference ordering $\succeq_\kappa$ based on the probability distribution κ has not the same meaning as the preference ordering $\succeq_\pi$ based on the possibility distribu-tion π. It is the same as saying that a probability distribution is not the same as a possibility distribution [10]. Their semantics are different. Therefore, passing from one model (or one ordering) to another via their respective normaliza-tion is not sensible. In other words, it is not meaningful to build a possibility

distribution from a probabilistic preference ordering; nor, conversely, to build a probability distribution from a possibilistic preference ordering.

2.6 Probability and non-specificity

First, the concept considered, of specificity for a probability distribution, is close to the specificity concept for a possibility distribution mentioned in section 2.3. Specificity is the ability, for a summative kernel or a probability distribution, to be concentrated on a set of minimal length (in the sense of the Lebesgue measure of this set). Measures of specificity should effectively capture the idea of how close a summative kernel is to a Dirac delta. The Dirac delta, at any given location ω of Ω, is the most specific probability distribution, whereas the uniform probability distribution on Ω, which equals $1/\lambda(\Omega)$ on Ω, should be the least specific distribution. λ is the Lebesgue measure defined on the Borel subsets of Ω. The first part of this remark, associated with theorem 5, leads to noting that for a family of summative kernels $(\kappa_\Delta^m)_{\Delta>0}$, given for any $m \in \Omega$, the specificity of κ_Δ^m increases when $\Delta \longrightarrow 0$, since this family tends to the Dirac delta.

We aim at defining, for summative kernels, an index of non-specificity reflecting its non-resolution power, in a way that is as natural as cardinality is an index of non-specificity for a maxitive kernel, reflecting its non-resolution power. As mentioned in the introduction, no such index has been defined in this manner in the literature. However, since a summative kernel is a probability distribution, classical probability distribution information indices could be used as non-specificity indices. In 1948, Shannon introduced a measure of information for probability distribution known as entropy [28]. For a probability distribution κ defined on a infinite set of alternatives Ω, the Shannon entropy is defined by:

$$H(\kappa) = -\int_\Omega \kappa(\omega) \log(\kappa(\omega)) d\omega, \tag{12}$$

with the convention $0 \log 0 := 0$.

This index is the unique solution of a set of axioms that may be found in [17], including the axioms of sub-additivity and expansibility. The axiom of sub-additivity says that the amount of information in a joint probability distribution cannot be greater than the sum of the amounts of information in the associated marginal probabilities. The expansibility axiom claims that expanding the set of alternatives with another alternative not supported by evidence, i.e. with probability 0, must not affect the amount of information. Rényi [23, 24] was probably the first to consider natural modifications in postulates of Shannon leading to non-Shannon entropies. The paper of Morales

et al. [18] is a very useful survey of such non-Shannon indices.

Sometimes the variance of (a random variable associated with) a probability distribution is also considered as a measure of non-specificity. Variance, which measures the average distance between observations of the underlying uncertain phenomenon, is an index of dispersion for potential observations. For a probability distribution κ defined on a infinite set of alternatives Ω, the variance is defined by:

$$V(\kappa) = \int_\Omega \omega^2 \kappa(\omega) d\omega - \left(\int_\Omega \omega \kappa(\omega) d\omega \right)^2. \tag{13}$$

Here is a simple example to illustrate the fact that the variance does not measure the specificity.
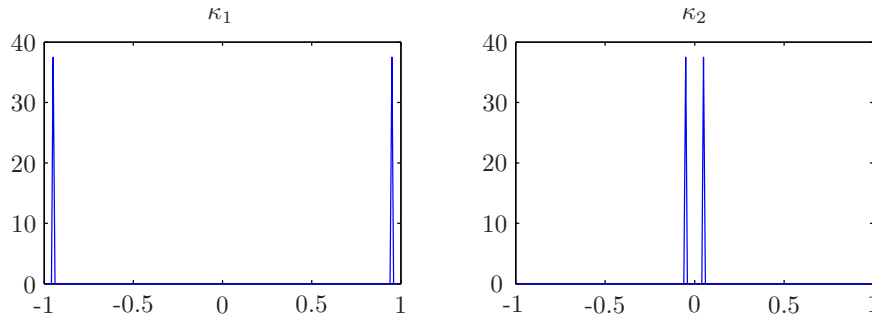
Figure 3. Probability distributions with same specificity

Figure 3 consists of two summative kernels, or probability distributions, κ_1 and κ_2. Each one is a mixture of two Epanechnikov distributions with $\Delta = 0.01$ defining the same weighted neighborhoods around two modes. For κ_1, the modes are $\{-0.95, 0.95\}$. For κ_2, the modes are $\{-0.05, 0.05\}$. Therefore, both κ_1 and κ_2 have concentration sets with the same length, since they both peak around two modes in the same way. Thus, they should have the same specificity index. The variance of κ_1 is greater than the variance of κ_2. According to this example, variance can not be considered as a good specificity index since it also reflects the average distance between the modes. More generally, variance measures the dispersion of potential observations of probability distributions.

3 Granulosity

3.1 Natural ordering of the alternatives based on confidence intervals

In the frequentist interpretation of probability, a confidence interval with confidence level α is an event (or a measurable set) of Ω such that the degree of

probability for any realization of the underlying uncertain phenomenon to fall in it equals α. In the framework of a subjective interpretation of probabilities, these intervals are named credible intervals. It is the key of the non-specificity index that we propose to capture the non-resolution power of a summative kernel. The specificity of a summative kernel can be locally observed for a
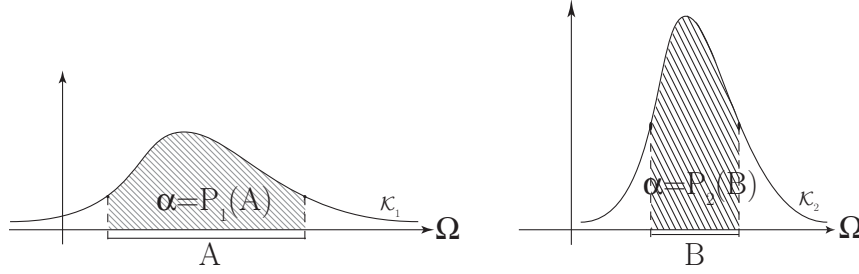

Figure 4. Local specificity comparison between κ_1 and κ_2

given degree. Figure 4 consists of two summative kernels κ_1 and κ_2, qualitatively drawn, such that κ_2 is more specific than κ_1. For a given level α, there are many confidence intervals. So as to compare the local specificity of two probability distributions, the smallest α confidence interval, in the sense of its Lebesgue measure, should be considered. This is called the most specific α confidence interval. It is A for κ_1 and B for κ_2. Since $\lambda(B) < \lambda(A)$, it seems natural to observe the specificity of a summative kernel, for a degree α, by looking at the length of its most specific α confidence interval. Therefore, a global index of specificity should aggregate the observed specificities over the degrees α.

In order to define this global specificity index, it will be easier to look at these confidence intervals in another direction, i.e. not by considering confidence intervals ranged by confidence level, but by looking at confidence intervals obtained from the elements ω of Ω. They are defined, for all ω of Ω, by

$$I_\omega = \{x \in \Omega / \kappa(x) \geq \kappa(\omega)\}. \qquad (14)$$

I_ω is a confidence interval with confidence level $P(I_\omega)$. The set of all intervals $(I_\omega)_{\omega \in \Omega}$ enables an ordering on the alternatives ω of Ω by significance, weight and contribution to measurement of the specificity of κ. Indeed, the more specific I_ω is (i.e. the smallest with respect to the Lebesgue measure), the more significant ω is in quantifying the specificity of κ. Figure 5 qualitatively illustrates this proposition. Each figure shows two confidence intervals, I_{ω_1} and I_{ω_2}, of the summative kernel κ seen as a probability distribution. I_{ω_2} is more specific than I_{ω_1}, therefore ω_2 should be more significant in the computation of the specificity index of κ than ω_1. Note that $(I_\omega)_{\omega \in \Omega}$ are nested sets. Indeed, for any ω_1 and ω_2 of Ω, such that $\kappa(\omega_1) \geq \kappa(\omega_2)$, we have $I_{\omega_1} \subseteq I_{\omega_2}$, by definition (14). Therefore, if I_{ω_1} is more specific than I_{ω_2}, then $\lambda(I_{\omega_1}) < \lambda(I_{\omega_2})$, then $I_{\omega_1} \subseteq I_{\omega_2}$, and therefore $P(I_{\omega_1}) < P(I_{\omega_2})$. This remark leads to a natural
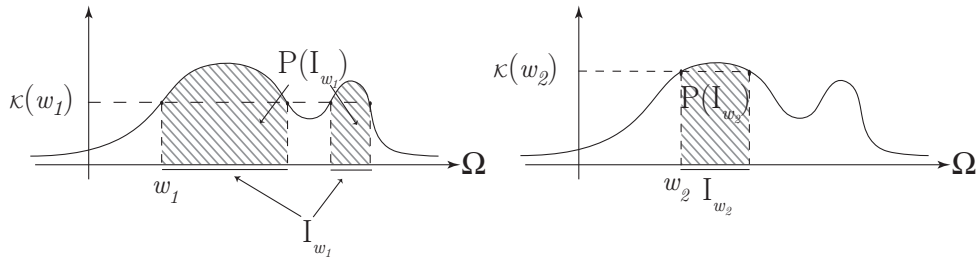
11

Figure 5. Specificity contribution of ω_1 and ω_2

preorder on the alternatives ω of Ω, according to their significance, for computing a specificity index of κ. This ordering, that we will note $\preceq_P$, is given by

$$\omega_1 \preceq_{sp} \omega_2 \iff P(I_{\omega_1}) \geq P(I_{\omega_2}).$$

It is equivalent to

$$\omega_1 \preceq_{sp} \omega_2 \iff 1 - P(I_{\omega_1}) \leq 1 - P(I_{\omega_2}). \tag{15}$$

The $\preceq_{sp}$ preorder on the alternatives of Ω is based on probability distribution κ but is different from $\preceq_\kappa$. Coupled with the construction procedure of a possibility distribution based on an order as summarized by expression (5), $\preceq_{sp}$ naturally leads to the construction of a possibility distribution $\pi_{\leftarrow\kappa}$ from κ, which has for possibility degrees:

$$\forall \omega \in \Omega, \ \pi_{\leftarrow\kappa}(\omega) = 1 - P(I_\omega), \tag{16}$$

since $1 - P(I_\omega) \in [0, 1]$, $\forall \omega \in \Omega$. In a more formal way, $\preceq_P \Leftrightarrow \preceq_{\pi_{\leftarrow\kappa}}$.

Obtaining this possibility distribution $\pi_{\leftarrow\kappa}$ is only one step in the definition of granulosity. $\pi_{\leftarrow\kappa}$ is the distribution which rearranges information concerning the specificity weights of the alternatives of Ω in an ordered way. $\pi_{\leftarrow\kappa}$ will be called the *possibility distribution for specificity of κ*.

Note that the possibility distribution obtained is exactly the solution of the probability/possibility transformation of Dubois and Prade [12, 6, 11].

3.2 Probability/possibility transformations

Dubois et al. have proposed three principles to be fulfilled by a sensible probability/possibility transformation: *consistency*, *preservation of the preference ordering* and *maximal specificity*.

Consistency between transformation components is an obvious first requirement for proper definitions of transformations between a probability distribution κ and a possibility distribution π. It postulates that any event $A \subseteq \Omega$

must have a higher degree of possibility than degree of probability, i.e. $\Pi(A) \geq P(A)$. It was first termed *consistency* by Zadeh [36], and is close to the notion of coherence in the imprecise probability theory of Peter Walley, where the domain Ω is discrete.

The *preservation of the preference ordering* can be defined in this way: an element ω_1 preferred to an element ω_2 in probability should be preferred in possibility. Formally, $\kappa(\omega_1) > \kappa(\omega_2) \Rightarrow \pi(\omega_1) > \pi(\omega_2)$ or $\omega_1 \succ_\kappa \omega_2 \Rightarrow \omega_1 \succ_\pi \omega_2$, when κ and π are respectively the probability and the possibility distribution associated with P and Π. Within this principle, equally probable elements do not need to be equally possible. Note that $\kappa(\omega_1) > \kappa(\omega_2) \Rightarrow I_{\omega_1} \subset I_{\omega_2} \Rightarrow 1 - P(I_{\omega_1}) > 1 - P(I_{\omega_2}) \Rightarrow \pi_{\leftarrow\kappa}(\omega_1) > \pi_{\leftarrow\kappa}(\omega_2)$.

When only the two previous principles are fulfilled, the transformation results in a set of possibility distributions instead of a unique possibility distribution.

Maximal specificity results in the choice of the possibility distribution having minimal cardinality. The possibility distribution fulfilling these principles is exactly the possibility distribution for specificity of κ, given by expression (16).

Dubois and Prade, in [6], proposed another transformation, that we note $\pi_{[\kappa]}$ and which is called the *subjective transformation*, because it is generally preferred to $\pi_{\leftarrow\kappa}$ in a subjective context. This is the converse transformation of the pignistic transformation defined by Smets [31, 13, 12]. It transforms a possibility measure Π into a probability measure P, which is the center of gravity of $\mathcal{M}(\Pi)$. The subjective transformation is defined here for a finite universe Ω by

Definition 6 *Let κ be a probability distribution defined on Ω. The possibility distribution $\pi_{[\kappa]}$, defined on Ω, obtained by the subjective transformation of κ, is defined by*

$$\forall \omega \in \Omega, \pi_{[\kappa]}(\omega) = \int_\Omega \min(\kappa(x), \kappa(\omega))dx. \tag{17}$$
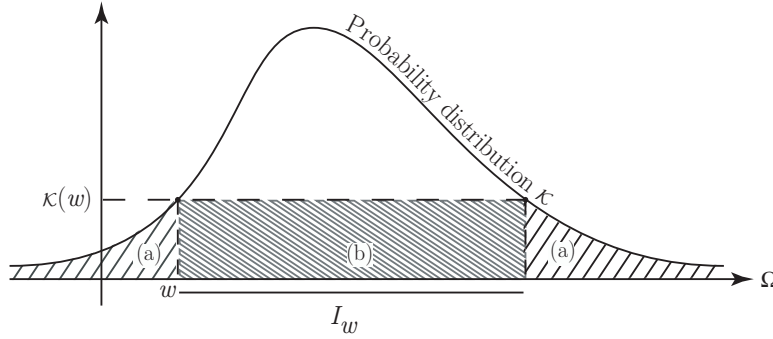

Figure 6. illustration of transformations $\pi_{[\kappa]}(\omega)$ and $\pi_{\leftarrow\kappa}(\omega)$

13

It is clear, when observing expression (17), that $\pi_{[\kappa]}(\omega)$ equals the hatched area $(a) + (b)$ of Figure 6 with $(a) = \pi_{\leftarrow\kappa}(\omega)$ and $(b) = \lambda(I_\omega)\kappa(\omega)$. The difference at ω between $\pi_{[\kappa]}(\omega)$ and $\pi_{\leftarrow\kappa}(\omega)$ is $(b) = \lambda(I_\omega)\kappa(\omega)$ and is positive. Thus, the transformation of Dubois and Prade results in a more specific possibility distribution than the subjective transformation, indeed $\forall \omega \in \Omega, \pi_{[\kappa]}(\omega) \geq \pi_{\leftarrow\kappa}(\omega)$.

Theorem 7 *Let κ be a probability distribution defined on Ω. The possibility distribution $\pi_{[\kappa]}$, d obtained by the subjective transformation, is given for every $\omega \in \Omega$, by*

$$\pi_{[\kappa]}(\omega) = \pi_{\leftarrow\kappa}(\omega) + \lambda(I_\omega)\kappa(\omega). \tag{18}$$

Table 1 is a non-exhaustive list of symmetric summative kernels, commonly used in signal processing applications, associated with their transformed maxitive kernels.

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

Table 1
common kernel transformations

The first column contains function u, which is the negative part of the basic summative kernel κ. While κ is even:

$$\kappa(\omega) = u(-|\omega|). \tag{19}$$

The second column contains function v, which is the negative part of the maxitive kernel $\pi_{\leftarrow\kappa}$. With κ being even, $\pi_{\leftarrow\kappa}$ is even and therefore:

$$\pi_{\leftarrow\kappa}(\omega) = v(-|\omega|). \tag{20}$$

The third column contains function t, which is the negative part of the maxitive kernel $\pi_{[\kappa]}$. With κ being even, $\pi_{[\kappa]}$ is even and therefore:

$$\pi_{[\kappa]}(\omega) = t(-|\omega|). \tag{21}$$

14

From Table 1, we can also retrieve κ_Δ and its associated maxitive kernels $\pi_{\leftarrow\kappa_\Delta}$ and $\pi_{[\kappa_\Delta]}$:

$$\kappa_\Delta(\omega) = \frac{1}{\Delta}u(-\frac{|\omega|}{\Delta}) = \frac{1}{\Delta}\kappa(\frac{\omega}{\Delta}), \tag{22}$$

$$\pi_{\leftarrow\kappa_\Delta}(\omega) = v(-|\frac{\omega}{\Delta}|) = \pi_{\leftarrow\kappa}(\frac{\omega}{\Delta}), \tag{23}$$

$$\pi_{[\kappa_\Delta]}(\omega) = t(-|\frac{\omega}{\Delta}|) = \pi_{[\kappa]}(\frac{\omega}{\Delta}). \tag{24}$$

The summative kernel, κ_Δ^m, whose mode is on ω_m, and its transformed maxitive kernels $\pi_{\leftarrow\kappa_\Delta^m}$ and $\pi_{[\kappa_\Delta^m]}$, are given by:

$$\kappa_\Delta^m(\omega) = \frac{1}{\Delta}u(-\frac{|\omega - \omega_m|}{\Delta}), \tag{25}$$

$$\pi_{\leftarrow\kappa_\Delta^m}(\omega) = v(-\frac{|\omega - \omega_m|}{\Delta}), \tag{26}$$

$$\pi_{[\kappa_\Delta^m]}(\omega) = t(-\frac{|\omega - \omega_m|}{\Delta}). \tag{27}$$

3.3 Granulosity

The key idea, in defining granulosity as a non-specificity index for a summative kernel κ, is the conjecture that the cardinality of $\pi_{\leftarrow\kappa}$ reflects the non-specificity of κ. In the converse sense, the specificity of probability distribution κ can be quantified by measuring specificity of $\pi_{\leftarrow\kappa}$.

In order to justify this conjecture, consider Figure 7. For a given level α, the alternatives x, with weights of specificity $\pi_{\leftarrow\kappa}(x)$ greater than α, form I_α, the α-cut of $\pi_{\leftarrow\kappa}$: $\{x \in \Omega/\pi_{\leftarrow\kappa}(x) \geq \alpha\}$. Because of the definition of $\pi_{\leftarrow\kappa}$, such a set can be identified with a set of the form given by expression (14): $\forall\alpha, \exists\omega/I_\alpha = I_\omega = \{x \in \Omega/\pi_{\leftarrow\kappa}(x) \geq \alpha\} = \{x \in \Omega/P(I_x) \leq 1 - \alpha\}$. The non-specificity index we propose is based on the idea that a summative kernel κ_2 is more specific than a summative kernel κ_1, if most of the $I_\alpha(\kappa_2)$ are smaller than the $I_\alpha(\kappa_1)$. Since the $I_\alpha(\kappa)$ are the α-cuts of $\pi_{\leftarrow\kappa}$, it is natural to measure the global specificity of κ by using a specificity measure of $\pi_{\leftarrow\kappa}$. This constructive characterization of the specificity of a summative kernel leads us to define our index of non-specificity, that we call granulosity, in the following way:

Definition 8 *Let κ be a summative kernel. The granulosity of κ, noted $\Gamma(\kappa)$, equals the cardinality of the possibility distribution $\pi_{\leftarrow\kappa}$:*

$$\Gamma(\kappa) = card(\pi_{\leftarrow\kappa}). \tag{28}$$

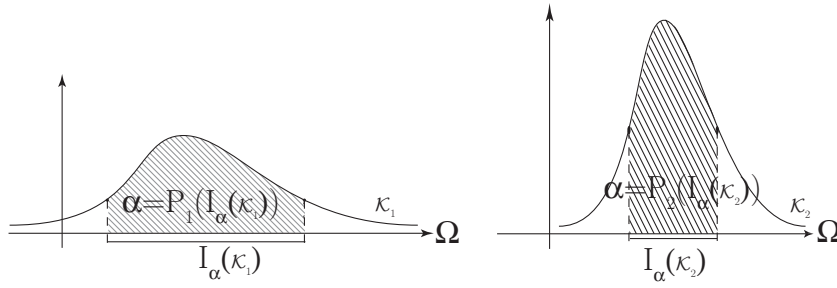This implies an ordering on probability distributions:

Figure 7.

Definition 9 *Let κ_1 and κ_2 be summative kernels defined on Ω, then κ_1 is said to be more granulated than κ_2 in the broad sense if and only if $\Gamma(\kappa_1) \geq \Gamma(\kappa_2)$, and strictly more granulated if and only if $\Gamma(\kappa_1) > \Gamma(\kappa_2)$.*

Note that this index is a continuous counterpart of an information index, studied by Dubois and Hüllermeier in [7] and proposed by Birnbaum [1] under the name of peakedness.

Beyond justifying the use of this index by intuitive considerations, Dubois and Hüllermeier in [7] proved theorem 10 that binds the peakedness index to the Shannon entropy of a probability vector $p = (p_i)_{i=1,..,|\Omega|}$. Let us define $\Lambda_\phi(.)$ as a generalized entropy given by

$$\Lambda_\phi(p) = \sum_{i=1}^{|\Omega|} \phi(p_i), \tag{29}$$

where the function $x \mapsto \phi(x)$ is strictly concave on $(0,1)$. Note that the function $x \mapsto -x \log(x)$ is strictly concave, which warrants the use of the term generalized entropy, since the Shannon entropy is defined by $H(p) = -\sum_{i=1}^{n} p_i \log(p_i)$,

Theorem 10 *If a probability vector p is less peaked than a probability vector q, then $\Lambda_\phi(p) \geq \Lambda_\phi(q)$, and if strictly less peaked, then $\Lambda_\phi(p) > \Lambda_\phi(q)$.*

In other words, the ordering based on the entropy is a refinement of the relative peakedness ordering.

These remarks, as transposed to continuous probabilities and restricted to the Shannon entropy, bring some meaning to the definition of granulosity. Indeed, it could be translated by the following conjecture:

Conjecture 11 *If a summative kernel κ_1 is more granulated than a summative kernel κ_2, then $H(\kappa_1) \geq H(\kappa_2)$, and, if strictly more granulated, then $H(\kappa_1) > H(\kappa_2)$.*

16

4 Adaptation of summative kernels in signal and image processing

4.1 Summative kernels in signal and image processing

In signal processing, one key procedure is the modeling of the acquisition of a real continuous signal by a sensor. One of the most common ways of performing this modeling is to associate a summative kernel with the impulse response of the sensor. As a converse problem, a wide range of signal analysis procedures inherently rely on the existence of methods for reconstructing a continuous signal from a set of uniformly sampled corrupted values obtained by acquiring a signal. This reconstruction procedure can also be achieved by the use of a summative kernel.

Most of the summative kernels used in signal processing are monomodal, symmetric, bounded and separable. For a summative kernel defined on an n-dimensional space Ω, separable means that it can be decomposed into a product of n summative kernels defined on the n marginal spaces. This is typically used to reduce the complexity of kernel-based algorithms. Table 2 offers a non-exhaustive list of such common kernels.

Sampling, acquisition modelling

Ideally, sensor acquisition at a location ω is modelled by a rectangular or crisp window which computes the average signal in a neighborhood of ω with radius Δ. Note that Δ can be considered as the bandwidth of an averaging summative kernel $r_\Delta = \frac{1}{\Delta}\mathbb{1}_{[-\Delta/2,\Delta/2]}$. Thus, r_Δ is assumed to be the impulse response of the sensor. The ideal measured signal, noted S and defined on Ω, is obtained by convolving the physical signal s with this averaging filter r_Δ,

$$S(\omega) = \int_\Omega s(u) r_\Delta(\omega - u) du = \frac{1}{\Delta} \int_{\omega-\Delta/2}^{\omega+\Delta/2} s(u) du. \tag{30}$$

This assumption of an averaging summative kernel is obviously naive, since such devices or sensors are seldom found in nature. Generally speaking, a sensor can be assimilated with an accumulator, that stores the information, or signal s, in a neighborhood of ω with radius Δ, thus providing the accumulation of the signal, i.e. a signal S proportional to the average of the signal s at ω over Δ. However, no real sensor works in this way, while physical accumulation generally involves capacity effects which tend to smooth its impulse response.

Thus, a relaxed or generalized model is usually preferred, consisting of replacing the averaging summative kernel r_Δ by a smoother summative kernel ν_Δ, which is more in accordance with sensor modeling. The role of the summative

kernel is to model the absorption of information contained in the underlying continuous physical signal around a given location. Then expression (30) becomes:

$$S(\omega) = \int_\Omega s(u)\nu_\Delta(\omega - u)du = (s * \nu_\Delta)(\omega). \qquad (31)$$

According to this model, a discrete signal can be seen as the values of the observed signal S at regularly (or possibly irregularly) distributed locations $(\omega_i)_{i=0,..,n}$ of Ω. We will restrict ourselves to the usual regularly distributed sampling, i.e. $\omega_i = \omega_0 + ih$, where h is the sampling step.

Definition 12 *A sampled signal consists of n measurements of an original continuous signal s at locations $(\omega_i)_{i=1,..,n}$, given by $S_i = S(\omega_i)$.*

Reconstruction, interpolation

Reconstruction can be presented as a reverse process, consisting of reconstructing a continuous signal, noted $\hat{S}$, from its discrete observations $(S_i)_{i=1,..,n}$.

Such a reconstruction mechanism can also be achieved by the use of summative kernels. In this case, the role of the summative kernel is to disperse or spread information contained in discrete observations, or measures, around them. Let η_Δ be the interpolation or reconstruction summative kernel with bandwidth Δ.

Definition 13 *Reconstruction of the continuous signal $\hat{S}$ from the sampled signal $(S_i)_{i=1,..,n}$, which is defined at the locations $(\omega_i)_{i=1,..,n}$ of Ω, is given, for all ω in Ω, by*

$$\hat{S}(\omega) = \sum_{i=1}^{n} \eta_\Delta(\omega - \omega_i)S_i = \left(\eta_\Delta * S\right)(\omega). \qquad (32)$$

This reconstruction is an interpolation if $\forall i \in \{1, ..., n\}, \ \hat{S}(\omega_i) = S_i$.

4.2   What is the adaptation of summative kernels?

The adaptation between two summative kernels $\kappa^1_{\Delta_1}$ and $\kappa^2_{\Delta_2}$, consists of objectively choosing the bandwidth Δ_2 of the kernel $\kappa^2_{\Delta_2}$, such that its behavior in classical kernel-based applications is as close as possible to the behavior of the summative kernel $\kappa^1_{\Delta_1}$ in the same application. It generally consists of finding a relation between κ^1, Δ_1 and κ^2, Δ_2, expressing a distance between the behaviors of $\kappa^1_{\Delta_1}$ and $\kappa^2_{\Delta_2}$, and deriving the adaptation relation from this relation, i.e.

$$\Delta_1 = f(\kappa^1, \kappa^2, \Delta_2) \ or \ \Delta_2 = g(\kappa^1, \kappa^2, \Delta_1). \qquad (33)$$

Two different approaches can be considered depending on the meaning of what is called the "behavior" of a summative kernel in a given application. The first

adaptation methodology consists of identifying behavior indices of the summative kernels. The second consists of optimizing the asymptotic behavior of Parzen Rosenblatt density estimators, which use the summative kernels involved in the adaptation. This simultaneous optimality gives rise to an explicit relation between the estimators, enabling the adaption of their asymptotic behavior.

4.3 Granulosity adaptation

The granulosity adaptation between two summative kernels κ_1 and κ_2 is based on the equalizing of their granulosities. The principle of this method is simple and intuitive: as the granulosity index of a kernel reflects its non-resolution power (i.e. its ability to collect or spread information), the equalizing of the granulosities of two kernels leads to a very natural adaptation of their behavior in applications where they can be seen as weighted neighborhoods ensuring smooth interplay between continuous and discrete domains.

The construction of an adaptation relation of the form (33), involving granulosity, requires a preliminary theorem. Let γ_κ be the granulosity of the basic summative kernel κ, i.e. $\gamma_\kappa = \Gamma(\kappa_1)$.

Theorem 14 *For any summative kernel κ_Δ,*

$$\Gamma(\kappa_\Delta) = \Delta\gamma_\kappa. \tag{34}$$

Proof 15 $\Gamma(\kappa_\Delta) = \int_\Omega \pi_{\leftarrow\kappa_\Delta}(\omega)d\omega = \int_\Omega \pi_{\leftarrow\kappa_1}(\frac{\omega}{\Delta})d\omega$, *see expression (23). Then,* $\Gamma(\kappa_\Delta) = \int_\Omega \Delta\pi_{\leftarrow\kappa_1}(\omega)d\omega = \Delta\gamma_\kappa.$ $\square$

As previously mentioned, we say that two kernels $\kappa^1_{\Delta_1}$ and $\kappa^2_{\Delta_2}$ are granulosity-adapted if $\Gamma(\kappa^1_{\Delta_1}) = \Gamma(\kappa^2_{\Delta_2})$. Because of theorem 14, this equation leads to $\Delta_1\gamma_{\kappa^1} = \Delta_2\gamma_{\kappa^2}$ and therefore:

$$\frac{\Delta_2}{\Delta_1} = \frac{\gamma_{\kappa^1}}{\gamma_{\kappa^2}} = \xi^{\kappa^2}_{\kappa^1}. \tag{35}$$

The granulosity adaptation relation is:

$$\Delta_2 = \xi^{\kappa^2}_{\kappa^1}\Delta_1, \tag{36}$$

$\xi^{\kappa^2}_{\kappa^1}$ is called the granulosity adaptation coefficient between κ^1 and κ^2. Kernel $\kappa^1_{\Delta_1}$ is said to be granulosity adapted to kernel $\kappa^2_{\Delta_2}$ if $\Delta_2 = \xi^{\kappa^2}_{\kappa^1}\Delta_1$.

The granulosity adaptation coefficient simply requires values of the basic granulosities γ_{κ^1} and γ_{κ^2} of the summative kernels κ^1 and κ^2. Table 2 offers a list of the granulosities of common basic summative kernels, thus facilitating the

19

recovery of the granulosity adaptation coefficient. For example, the adaptation coefficient between the triangular kernel T and the Epanechnikov kernel E is given by $\xi_T^E = \gamma_T/\gamma_E = 8/9$.

4.4 Entropy adaptation

Shannon entropy adaptation is based on the identification of the Shannon entropies of summative kernels κ^1 and κ^2 involved in the adaptation. This is exactly the same principle as that of granulosity adaptation.

The construction of an adaptation relation of the form (33), involving the Shannon entropy, requires a preliminary theorem. Let h_κ be the Shannon entropy of the basic summative kernel κ, i.e. $h_\kappa = H(\kappa_1)$.

Theorem 16 For any summative kernel κ_Δ,

$$H(\kappa_\Delta) = \log(\Delta) + h_\kappa. \tag{37}$$

Proof 17 $H(\kappa_\Delta) = - \int_\Omega \kappa_\Delta(\omega) \log(\kappa_\Delta(\omega)) d\omega = - \int_\Omega \frac{1}{\Delta}\kappa_1(\frac{\omega}{\Delta}) \log(\frac{1}{\Delta}\kappa_1(\frac{\omega}{\Delta})) d\omega = - \int_\Omega \kappa_1(x)(\log(\frac{1}{\Delta}) + \log(\kappa_1(x))) dx = \log(\Delta) \int_\Omega \kappa_1(x) dx - \int_\Omega \kappa_1(x) \log(\kappa_1(x)) dx = \log(\Delta) + h_\kappa. \ \square$

As previously mentioned, we say that two kernels $\kappa_{\Delta_1}^1$ and $\kappa_{\Delta_2}^2$ are entropy-adapted if $H(\kappa_{\Delta_1}^1) = H(\kappa_{\Delta_2}^2)$. Because of theorem 16, this equation leads to $\log(\Delta_1) + h_{\kappa^1} = \log(\Delta_2) + h_{\kappa^2}$ and therefore:

$$\frac{\Delta_2}{\Delta_1} = e^{h_{\kappa^1} - h_{\kappa^2}} = \phi_{\kappa^1}^{\kappa^2}. \tag{38}$$

The entropy adaptation relation is:

$$\Delta_2 = \phi_{\kappa^1}^{\kappa^2} \Delta_1, \tag{39}$$

$\phi_{\kappa^1}^{\kappa^2}$ is called the Shannon entropy adaptation coefficient between κ^1 and κ^2. Kernel $\kappa_{\Delta_1}^1$ is said to be entropy adapted to kernel $\kappa_{\Delta_2}^2$ if $\Delta_2 = \phi_{\kappa^1}^{\kappa^2} \Delta_1$.

The entropy adaptation coefficient simply requires values of the basic entropies h_{κ^1} and h_{κ^2} of the summative kernels κ^1 and κ^2. Table 2 offers a list of the entropies of common basic summative kernels, thus facilitating the recovery of the entropy adaptation coefficient. For example, the entropy adaptation coefficient between the triangular kernel T and the Epanechnikov kernel E is given by $\phi_T^E = e^{h_T - h_E} = 0.9343$.

20

The *AMISE* approach [30] belongs to the second family of adaptations. In fact, the ability of a summative kernel to collect information is used in the so-called Parzen Rosenblatt [20, 25] density estimator $f_{\kappa\Delta}$. The probability density function f, describing a random variable X on Ω, is estimated by using a sample of N independent and identically-distributed observations $(X_1, ..., X_N)$ of X, and a summative kernel κ_Δ.

$$f_{\kappa\Delta}(\omega) = \frac{1}{N\Delta} \sum_{i=1}^{N} \kappa(\frac{\omega - X_i}{\Delta}) = \frac{1}{N} \sum_{i=1}^{n} \kappa_\Delta(\omega - X_i). \tag{40}$$

The asymptotic behaviour of the Parzen Rosenblatt density estimator can be characterized by an upper bound of the L_2-error (also called the Mean Integrated Squared Error) between the density estimator $f_{\kappa\Delta}$ and the underlying density function f of the random variable X, given by $MISE = \|f - f_{\kappa\Delta}\|_{L_2} = \int_\Omega (f(\omega) - f_{\kappa\Delta}(\omega))^2 d\omega$. The upper bound of this L_2-error, called the Asymptotic Mean Integrated Squared Error: $AMISE$, in the case of the Parzen Rosenblatt estimator $f_{\kappa\Delta}$, is analytically given by : $AMISE(f_{\kappa\Delta}) = \frac{R(\kappa)}{N\Delta} + \frac{\Delta^4 \sigma_\kappa^4 R(f'')}{4}$, with $R(\phi) = \int \phi^2(u)du$ and $\sigma_\phi^2 = \int u^2 \phi(u)du$.

Next, as proved in [22, 27, 29], the bandwidth, Δ^*, which minimizes the $AMISE$ is given by:

$$\Delta^* = \Big(\frac{R(\kappa)}{\sigma_\kappa^4}\Big)^{\frac{1}{5}} \Big(R(f'')N\Big)^{-\frac{1}{5}}. \tag{41}$$

Note that there is no one-to-one correspondence between κ and its associated optimal bandwidth Δ^*. Indeed, Δ^* depends not only on f, the density to estimate and N, the number of observations, but also on κ, the kernel chosen for the estimation.

The *AMISE* adaptation involves computation of the ratio of the two optimal bandwidths Δ_1^* and Δ_2^*, obtained in the same estimation conditions (i.e. the same underlying function f and the same number of observations N) from κ^1 and κ^2. This ratio is noted $\zeta_{\kappa^1}^{\kappa^2}$:

$$\zeta_{\kappa^1}^{\kappa^2} = \frac{\Delta_2^*}{\Delta_1^*} = \Big(\frac{R(\kappa^2)}{\sigma_{\kappa^2}^4}\Big)^{\frac{1}{5}} \Big(\frac{R(\kappa^1)}{\sigma_{\kappa^1}^4}\Big)^{-\frac{1}{5}}, \tag{42}$$

which depends neither on f nor on N.

The *AMISE* adaptation relation is:

$$\Delta_2 = \zeta_{\kappa^1}^{\kappa^2} \Delta_1. \tag{43}$$

$\zeta_{\kappa^1}^{\kappa^2}$ is called the *AMISE* adaptation coefficient between κ^1 and κ^2. Kernel $\kappa_{\Delta_1}^1$ is said to be *AMISE*-adapted to kernel $\kappa_{\Delta_2}^2$ if $\Delta_2 = \zeta_{\kappa^1}^{\kappa^2}\Delta_1$.

This is, however, a shortcoming of this method which is applied to the whole bandwidth range, whereas it is fully valid only for the optimal bandwidth.

4.6 Indices and adaptation coefficients comparisons

Table 2 presents a comparison of the granulosity index γ_κ and the Shannon entropy index h_κ for nine common continuous summative kernels κ. The entropy indices are sorted by increasing order. This order is the same as the granulosity index order. This suggests that theorem 10 is also verified for continuous probability distributions.

	$\kappa(\omega)$	h_κ	γ_κ
triweight	$\frac{35}{32}(1-\omega^2)^3$	0.3086	0.5469
triangular over quadratic	$\frac{1}{\pi/2-\ln 2}\left(\frac{1-|\omega|}{1+\omega^2}\right)$	0.4220	0.6015
triangular	$1-|\omega|$	$1/2$	0.6667
cosine	$\frac{\pi}{4}\cos(\frac{\pi}{2}\omega)$	0.5484	0.7268
Epanechnikov	$\frac{3}{4}(1-\omega^2)$	0.5680	0.75
trapezial	$\frac{2}{3}\mathbb{1}_{|\omega|<\frac{1}{2}} + \frac{4}{3}(1-|\omega|)\mathbb{1}_{|\omega|\geq\frac{1}{2}}$	0.5721	0.7778
uniform	$\frac{1}{2}$	0.6931	1
gaussian	$\frac{1}{\sqrt{2\pi}}e^{-\frac{\omega^2}{2}}$	1.4189	1.5948
exponential	$\frac{1}{2}e^{-|\omega|}$	1.6931	2

Table 2
Comparison between the entropy and the granulosity of common kernels

Table 3 presents the adaptation coefficients of three summative kernels to the uniform kernel. The second column presents the *AMISE* adaptation coefficients [30], computed with expression (42). The third column presents the granulosity adaptation coefficients computed with expression (35). The fourth column presents the entropy adaptation coefficients computed with expression (38). The three adaptation coefficients are quite close and seem to go in the same direction, i.e. $\zeta_U^{\kappa^1} > \zeta_U^{\kappa^2} \Leftrightarrow \xi_U^{\kappa^1} > \xi_U^{\kappa^2} \Leftrightarrow \phi_U^{\kappa^1} > \phi_U^{\kappa^2}$. In the next section, we propose experiments based on these adaptations to highlight the fact that granulosity properly measures the non-resolution power of the summative kernels used in common signal processing applications.

	granulosity ξ_U^κ	entropy ϕ_U^κ	$AMISE$ ζ_U^κ
Epanechnikov: $\frac{3}{4}(1-x^2)\mathbb{1}_{[-1,1]}$	1.3333	1.1333	1.2724
Triweight: $\frac{35}{32}(1-x^2)^3\mathbb{1}_{[-1,1]}$	1.8286	1.4689	1.7115
Gaussian: $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	0.6270	0.4839	0.5747

Table 3

Comparison between $AMISE$, granulosity and entropy adaptation coefficients

5 Experiments

We carried out numerous experiments on applications using the summative kernels: filtering, impulse response modeling, image decimation. We present here some results concerning image filtering and sensor modeling.

5.1 Experimental comparison between adaptation methods in signal filtering

This experiment concerns adaptation comparison in a filtering context. As an illustrative experiment, we propose to filter a 3200×3200 satellite image using 4 different summative kernels: Gaussian, Uniform, Epanechnikov and Triweight. This satellite image, depicted in Figure 8, was provided by the Institut Géographique National (IGN) and represents a part of the Bouches-du-Rhône. The sampling step is taken as a unit, i.e. $h = 1$.

Figure 8. Bouches-du-Rhône

The images resulting from the filtering are compared when the bandwidths of

the 4 summative kernels are adapted with $AMISE$, entropy and granulosity adaptations. The comparison has also been effected with no adaptation. Figure 9 shows the mean of the L_2-distances between the filtered images for different values of the bandwidth of the uniform kernel. The values of the bandwidths of the other summative kernels are obtained using the adaptation relations (36), (39) and (43). The comparison was performed on the central part of the filtered images to avoid side effects.

As a preliminary trivial remark, any adaptation is better than no adaptation, whatever the chosen method. Using the worst adaptation method will, on average, provide a mean L_2-distance that is 20 times smaller than the mean L_2-distance obtained without adaptation. This result is not shown on figure 9, because it could lower the informativity of the plotting.
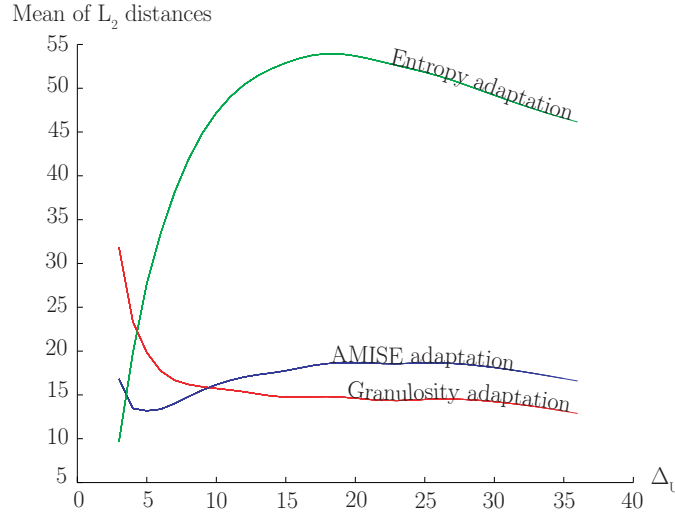

Figure 9. Efficiency comparison between adaptation methods in image filtering

When the bandwidths are quite small ($\Delta_U \leq 10$), the filtered images are not smooth enough for the filters to provide equivalent results (see Figures 10, 11 and 12). In this case, adaptation with $AMISE$ provides a better result (i.e. the images are closer) than granulosity adaptation. Conversely, when the adapted bandwidth exceeds this value, the filtering is strong enough to clear the main details from the satellite image (see Figures 13, 14 and 15), and the dissimilarities between the different results are stabilized. In this last case, granulosity adaptation provides a better result than $AMISE$ adaptation. Therefore, since filtering consists of removing details, granulosity adaptation seems to be more appropriate in this kind of application.

Now, when comparing granulosity adaptation to Shannon adaptation, the same remarks are valid with a lower threshold ($\Delta_U < 4$). Therefore, granulosity adaptation seems to be more appropriate than entropy adaptation in image filtering. Actually, entropy adaptation does not seem to be a proper adaptation

method in such a context: the distance between the filtered adapted images tends to increase as Δ increases.

Note that, in this kind of discrete filtering application, the spreading and collecting properties of a summative kernel can hardly be differentiated. In fact, the filtered value of a discrete signal at a given location with a given kernel can be thought of either as the spreading, with this kernel, of all the sampled values evaluated at the considered location, or as the collecting of all sampled values in the weighted neighborhood defined by the kernel at the considered location. Therefore, regarding filtering application, the spreading and collecting abilities of a given summative kernel can be sought as dual properties.

In this kind of application, the number of sampled values that the filter absorbs or spreads depends on the bandwidth of the summative kernel. Remember that the sampling step is $h = 1$. A uniform kernel-based filter, with a bandwidth of 5, will act on 11×11 pixels. More generally, a uniform kernel-based filter, with a bandwidth Δ_U, will act on $2\Delta_U + 1 \times 2\Delta_U + 1$ pixels. Therefore, the uniform filter acts on sets of pixels whose cardinality increases with the bandwidth. This property is also consistent with other summative kernels. Since, for large bandwidths, granulosity adaptation is better than other adaptation methods, the granulosity index can be seen as a good marker of the absorption or spreading abilities of a summative kernel, and therefore is a good index of non-resolution power.

Figure 10. $\Delta_U = 1$ Figure 11. $\Delta_U = 5$ Figure 12. $\Delta_U = 10$

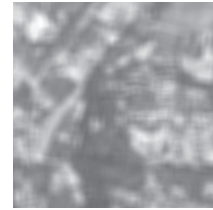Figure 13. $\Delta_U = 20$ Figure 14. $\Delta_U = 25$ Figure 15. $\Delta_U = 30$

Figures 10 to 15 represent filtered images of the satellite image depicted on figure 8, obtained with the uniform kernel whose bandwidth is taken from $\Delta_U = 1$ to $\Delta_U = 30$.

The property we aim to highlight, within this experiment, is the ability of the granulosity index to characterize the absorption capability of a summative kernel. This experiment consists of simulating the behavior of one detector of a nuclear imaging device [14]. Such a detector is designed to count the photons emitted in a certain direction during a given period. A complete device is made up of several detectors. Each one is associated with a collecting direction. The proportion of photons detected by a sensor over the complete number of detected photons is the signature of the density of radioactivity in the direction associated to the sensor. The density of radioactivity detected in a certain direction is known to be ruled by a Poisson process, and depends not only on the measured density of radioactivity but also on N, the total number of detected photons, and on κ, the impulse response of the sensor.

In this experiment, the radioactive zone is supposed to be approximately punctual, which means that it is assumed to be symmetrically distributed. We perform this experiment by modeling this radioactive zone by a centered uniform distribution on $[-50, 50]$. We suppose that all the photons emitted by this active zone are detected by the device, and we focus on the density of photons detected by one of the detectors.

We compute the detected density when modeling the impulse response of this sensor by the four summative kernels used in the previous experiment: Uniform, Epanechnikov, Triweight and Gaussian, with adapted bandwidths. This computation is made for Δ_U ranging from 0 to 10 and for a number of emitted photons ranging from 1 to 1000.

We compute the L_2-distances between the densities obtained with the four presented summative kernels when using different adaptation methods and, also, with no adaptation. These L_2-distances are computed over 100 different experimental data sets.
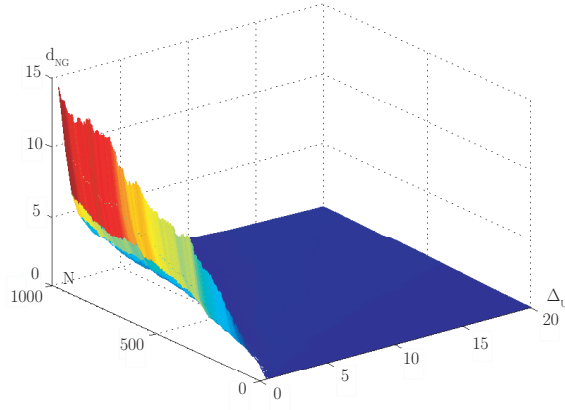
Figure 16. Comparison between no adaptation and granulosity adaptations

As a first remark, whatever Δ_U and N, granulosity adaptation is better than no adaptation. Figure 16 shows the difference, noted d_{NG}, between the average distance with no adaptation and the average distance with granulosity adaptation. The results obtained for the other adaptation methods are similar, and therefore not plotted here.
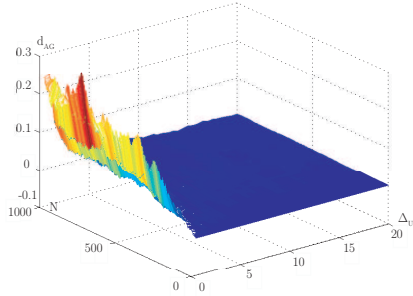

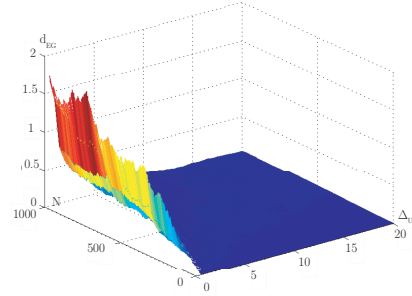Figure 17. Comparison between AMISE and granulosity adaptations

Figure 18. Comparison between entropy and granulosity adaptations

Figure 17 shows the difference, noted d_{AG}, between the average distance with AMISE adaptation and the average distance with granulosity adaptation. For most of the parameters Δ_U and N, the granulosity adaptation is better than the AMISE adaptation method, except for some pairs (N, Δ_U), with small N and small Δ_U.

Figure 18 shows the difference, noted d_{EG}, between the average distance with entropy adaptation and the average distance with granulosity adaptation. Whatever Δ_U and N, the granulosity adaptation is better than the entropy adaptation method. Since this experiment uses the absorption property of the

summative kernels, we can conclude that the granulosity better captures the ability of a summative kernel to absorb information than the entropy does.

6   Conclusion

In a digital signal processing context, summative kernels are widely used either to model the impulse response of a sensor, or to ensure a smooth interplay between continuous and discrete domains. In such a context, it is of practical importance to be able to use an index that characterizes the behavior of a summative kernel in different applications.

When a kernel is used to model the impulse response of a sensor, one of the characteristics of practical importance to be derived from the kernel is the resolution power of the sensor. This resolution power is usually defined as being inversely proportional to the minimal distance between two measured features that can be separated by using the measured signal. When modeling the measurement by a uniform kernel defined on an interval, this minimal distance is directly proportional to the length of the interval. Since this length is, by definition, the granulosity of a uniform kernel, it seems straightforward to conjecture that the granulosity of any summative kernel associated with the impulse response of a sensor is a marker of its non-resolution power. Granulosity can thus be seen as an index reflecting non-resolution power, i.e. the ability of a summative kernel to spread or collect information in the usual signal processing applications. Note that the Shannon entropy index is not consistent with this definition of a non-resolution power index, since the Shannon entropy index of a uniform kernel defined on an interval tends to $-\infty$ as its length tends to 0.

A key step in our definition of granulosity is to conjecture that the (non-)resolution power of a summative kernel and the (non-)specificity of a probability distribution are dual properties. In fact, we define the granulosity of a summative kernel as an index characterizing its non-specificity, considering this summative kernel as being a probability distribution. This specificity is the ability, for a probability distribution, to be concentrated on a set of minimal Lebesgue measure. Birnbaum [1], by means of the peakedness index, defined a non-specificity measure that has been studied by Dubois and Hüllermeier in [7]. It has been observed that the construction of the peakedness is based on the probability/possibility transformation of Dubois and Prade [12]. It appears that granulosity is simply the peakedness index of Birnbaum, defined for continuous probability distributions.

The numerous experiments we carried on highlighted the fact that, when two summative kernels with different shapes have the same granulosity, their be-

28

havior in typical kernel-based signal processing applications are quite close, and, moreover, closer than two summative kernels having the same Shannon index. For adaptation purposes, the AMISE adaptation method derived from the Parzen-Rosenblatt kernel-based density estimation may also be considered. In fact, in applications where few data are involved, this adaptation sometimes seems to work better than the granulosity-based adaptation. However, even if the AMISE adaptation can be useful in comparing summative kernels, it seems difficult to derive a direct, useful, non-resolution index from the adaptation coefficients. It should still be an interesting research track to follow.

Among numerous possible uses, our new index could also yield new criteria for finding the optimal shape of a summative kernel for a particular application. Moreover, due to the separability property of theorem 14, the use of such criteria could lead to very simple methods, and to computationally low-cost algorithms. One could also consider directly infering the non-resolution power of a sensor by applying the probability/possibility transformation directly to the experimental data collected when the sensor is measuring a known pattern. Finally, since a maxitive kernel can be viewed as representing the family of the summative kernels with lower or equal granulosity, it sounds sensible to consider a new way of performing signal processing based on replacing summative kernels by maxitive kernels in the usual applications.

References

[1] Z. W. Birnbaum. On random variables with comparable peakedness. *Annals of Mathematical Statistics*, 19:76–81, 1948.

[2] S. Bodjanova. Granulation of a fuzzy set: Nonspecificity. *Information Sciences*, 177:4430–4444, 2007.

[3] J. Canny. A computational approach to edge detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 8, pages 679–698, 1986.

[4] G. De Cooman and D. Aeyels. Supremum-preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.

[5] 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.

[6] D. Dubois. Possibility theory and statistical reasonning. *Computational Statistics and Data Analysis*, 51:47–69, 2006.

[7] D. Dubois and E. Hüllermeier. Comparing probability measures using possibility theory: a notion of relative peakedness. *International Journal of Approximate Reasoning*, 2006.

[8] D. Dubois and H. Prade. *Théorie des possibilités : Applications à la représentation des connaissances en informatique*. Masson, 1988.

[9] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.

[10] D. Dubois and H. Prade. Fuzzy sets and probability: misunderstandings, bridges and gaps. In *Proceedings of the Second IEEE Conference on Fuzzy Systems*, pages 1059–1068, 1993.

[11] D. Dubois, H. Prade, L. Foulloy, and G. Mauris. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10:273–297, 2004.

[12] 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.

[13] 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.

[14] 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.

[15] F. Jacquey, K. Loquin, F. Comby, and O. Strauss. Non-additive approach for gradient-based edge detection. In *ICIP07, International Conference on Image Processing*, 2007.

[16] Jiri Jan. *Digital Signal Filtering, Analysis and Restoration*. IET, 2000.

[17] G. J. Klir and M.J. Wierman. *Uncertainty Based Information: Elements of Generalized Information Theory*. Physica Verlag, 1998.

[18] 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.

[19] H. T. Nguyen. Some mathematical tools for linguistic probabilities. *Fuzzy Sets and Systems*, 2:53–65, 1979.

[20] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[21] Z. Pawlak. *Rough sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publisher, 1991.

[22] B. L. S. Prakasa Rao. *Nonparametric functional estimation*. Academic Press, 1983.

[23] A. Rényi. On the measures of entropy and information. In *4th Berkeley Symp. Math. Statist. and Prob.*, volume 1, pages 547–561, 1961.

[24] A. Rényi. On the foundations of information theory. *Rev. Int. Statist. Inst.*, 33:1–14, 1965.

[25] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27:832–837, 1956.

[26] L. Schwartz. *Théorie des distributions*, volume 1. Hermann, Paris, 1950.

[27] D. W. Scott. *Multivariate Density Estimation*. Wiley Interscience, 1992.

[28] C. E. Shannon. A mathematical theory of communication. *The Bell System technical journal*, 27:379–423, 1948.

[29] B. W. Silvermann. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability 26, Chapman and Hall, 1986.

[30] J. S. Simonoff. *Smoothing Methods in Statistics*. Springer-Verlag, 1996.

[31] Ph. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

[32] M. Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Process. Mag.*, 16:22–38, 1999.

[33] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.

[34] P. Walley. Towards a unified theory of imprecise probability. *Int. J. Approx. Reasoning*, 24(2-3):125–148, 2000.

[35] R. R. Yager. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9:249–260, 1983.

[36] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.