WCCI 2010 IEEE World Congress on Computational Intelligence
July, 18-23, 2010 - CCIB, Barcelona, Spain FUZZ-IEEE

Towards an interval-valued estimation of the density

Bilal Nehme and Olivier Strauss

***Abstract*— This paper presents a theoretical and practical novel approach for computing the probability density function underlying a set of observations. The estimator we propose is an extension of the conventional Parzen Rosenblatt method that leads to a very specific interval-valued estimation of the density. Within this approach, we make use of the convenient representation of a set of usual (summative) kernels by a maxitive kernel (i.e. a possibility distribution) to derive an exact computation with a very low complexity of an interval-valued estimation. The considered set of kernels is particularly convenient since it contains kernels having comparable shapes and bandwidth. We prove that the obtained imprecise probability density function contains a set of precise density functions estimated using the standard method with kernels belonging to the considered set.**

I. Introduction

In the last ten years, there has been an increasing interest for imprecise probability in decision making. This framework has been mainly developed in the context of epistemic probabilities to handle the difficulty of representing ill known information by mean of precise subjective probabilities.

In the more objective context of real data processing, this framework is barely used due to the difficulty of computing or specifying the imprecise probability of any event associated to an observation process (e.g. measurement, sensing, imaging, etc.). Yet, such a knowledge could lead to new robust method for analyzing, filtering or comparing data.

In this context of real data processing, probability density function (pdf) plays a central role. Assuming a particular form of this density able, for example, comparing two sets of data via a particular distance (e.g. Mahalanobis distance) or filtering the data to remove spurious random variations. If not, this density has to be estimated from a finite sample of observations supposedly independent and identically distributed.
There are two classical ways to achieve such an estimation: parametric and nonparametric, depending on whether or not a particular model can be assumed for the density [10], [11]. Here, we concentrate on nonparametric approaches. Among the different non parametric approaches [12], the most popular is the so-called Parzen Rosenblatt method. This popularity comes from its easy computation and the easy interpretation of the density it provides [13].

In this paper, we propose an extension of the Parzen Rosenblatt approach which leads to an interval-valued probability density function [2]. Due to its construction, this interval-valued probability has a particular meaning: it is the convex set of all Parzen Rosenblatt estimations obtained with kernel having comparable shapes and comparable bandwidths. This work is restricted to symmetric bounded kernel having a first derivative, i.e. the kernel that are mostly used in this context. We also propose a practical, exact and low cost implementation of the computation of this interval-valued pdf.

This paper is organized as follows: section II introduces some necessary preliminary concepts and reformulates the Parzen Rosenblatt estimator in a way that can lead to our interval-valued estimator. In section III we show how to built our interval-valued estimator based on a particular kernel. Section IV presents the discrete algorithm leading to an exact computation of the interval valued probability in any point of a reference interval. Section V illustrates the proposed interval-valued estimation.

Bilal Nehme is with the Department of Robotics, LIRMM, 161, rue Ada, 34095 Montpellier Cedex 5, France (phone: +33 467 418 564; email: Bilal.Nehme@lirmm.fr).
Olivier Strauss is with the Department of Robotics, LIRMM, 161, rue Ada, 34095 Montpellier Cedex 5, France (phone: +33 467 418 587; email: Olivier.Strauss@lirmm.fr).

II. Preliminaries

This preliminary section aims at presenting different mathematical tools that will be used to construct the imprecise pdf estimator we propose. In the rest of the paper Ω will be the interval $[e_{min}, e_{max}]$ of $\mathbb{R}, \mathcal{P}(\Omega)$ the collection of all Lebesgue measurable subsets of Ω, $\mathcal{K}(\Omega)$ the set of summative kernels in Ω and $s : \Omega \to \mathbb{R}$ a bounded L_1 function associated to a distribution in the meaning of Schwartz [9].

A. Summative and maxitive kernel

Kernels are functions from Ω to $\mathbb{R}$ that are often used in signal processing and non-parametric statistics to defining a weighted neighborhood around a location $u \in \Omega$. The distinction between summative and maxitive kernels has been introduced by [4] to handle an imprecise knowledge on the proper tool to be used in a data processing context.

A **summative kernel** is a positive real valued function κ of Ω, verifying the summativity property:

$$\int_\Omega \kappa(u)du = 1. \quad (1)$$

A summative kernel κ can be seen as a probability distribution inducing an additive confidence measure (probability) P_κ defined by:

$$\forall A \in \mathcal{P}(\Omega), P_\kappa(A) = \int_A \kappa(u)du.$$

A summative kernel κ_Δ^x can be defined by translating a generic summative kernel κ in $x \in \Omega$ by:

$$\forall u \in \Omega, \kappa_\Delta^x(u) = \frac{1}{\Delta}\kappa(\frac{u-x}{\Delta}),$$

with $\Delta > 0$. Δ is called the bandwidth of κ_Δ^x. By construction, $\forall u \in \Omega,\ \kappa(u) = \kappa_1^0(u)$.

A **maxitive kernel** is a real function π from Ω to $[0,1]$, verifying the maxitivity property:

$$\sup_{u\in\Omega} \pi(u) = 1. \tag{2}$$

A maxitive kernel can be seen as a possibility distribution, inducing two dual non-additive confidence measures on Ω: a possibility measure Π_π and a necessity measure N_π, defined by:

$$\forall A \in \mathcal{P}(\Omega), \Pi_\pi(A) = \sup_{x\in A} \pi(x) \quad (possibility),$$

$$N_\pi(A) = 1 - \Pi_\pi(A^c) \quad (necessity),$$

with A^c being the complementary set of Ω.
A maxitive kernel π_Δ^x can be defined by translating a generic maxitive kernel π in $x \in \Omega$ by:

$$\forall u \in \Omega, \pi_\Delta^x(u) = \pi(\frac{u-x}{\Delta}),$$

with $\Delta > 0$. Δ is called the bandwidth of π_Δ^x. By construction, $\forall u \in \Omega,\ \pi(u) = \pi_1^0(u)$.

A maxitive kernel π is said to dominate a summative kernel κ [4] if the possibility measure Π_π dominates the probability measure P_κ, i.e.:

$$\forall A \in \mathcal{P}(\Omega), P_\kappa(A) \leq \Pi_\pi(A).$$

In that sense, a maxitive kernel defines a set of summative kernels denoted $\mathcal{M}(\pi)$ and defined by:

$$\mathcal{M}(\pi) = \begin{Bmatrix} \kappa \in \mathcal{K}(\Omega), \text{such that} \\ \forall A \in \mathcal{P}(\Omega), N_\pi(A) \leq P_\kappa(A) \leq \Pi_\pi(A) \end{Bmatrix}.$$

B. Derivative of a symmetric summative kernel

In most data processing applications, the kernel used are unimodal and symmetric with a bounded support and having a first derivative. The symmetry will be an important property that will be used in the construction of our new estimator. In the rest of the paper $\mathcal{K}'(\Omega)$ will denoted the subset of unimodal symmetric kernels of $\mathcal{K}(\Omega)$ having a bounded support and a first derivative.

Let $\kappa_\Delta \in \mathcal{K}'(\Omega)$. The Jordan decomposition of its first derivative $d\kappa_\Delta$ is given by: $-d\kappa_\Delta = d\kappa_\Delta^+ - d\kappa_\Delta^-$, with $d\kappa_\Delta^+ = max(0, -d\kappa_\Delta)$ and $d\kappa_\Delta^- = max(0, d\kappa_\Delta)$.

Property 1: Let $\kappa_\Delta \in \mathcal{K}'(\Omega)$. The derivative of a summative kernel κ_Δ can be written as the linear combination of two summative kernels η_Δ^+ and η_Δ^-:

$$\forall u \in \Omega, -d\kappa_\Delta(u) = a_\Delta^+ \eta_\Delta^+(u) - a_\Delta^- \eta_\Delta^-(u), \tag{3}$$

where a_Δ^+ and a_Δ^- are two constants defined by $a_\Delta^+ = \int_\Omega d\kappa_\Delta^+(u)du$ and $a_\Delta^- = \int_\Omega d\kappa_\Delta^-(u)du$.

Proof η_Δ^+ and η_Δ^- are defined by: $\eta_\Delta^+(u) = \frac{d\kappa_\Delta^+(u)}{a_\Delta^+}$ and $\eta_\Delta^-(u) = \frac{d\kappa_\Delta^-(u)}{a_\Delta^-}$.
η_Δ^+ and η_Δ^- are positive by construction and follow the summativity condition:

$$\int_\Omega \eta_\Delta^+(u)du = \frac{1}{a_\Delta^+}\int_\Omega d\kappa_\Delta^+(u)du = 1,$$

and

$$\int_\Omega \eta_\Delta^-(u)du = \frac{1}{a_\Delta^-}\int_\Omega d\kappa_\Delta^-(u)du = 1.$$

Since, by construction $d\kappa_\Delta^+ = a_\Delta^+\eta_\Delta^+$ and $d\kappa_\Delta^- = a_\Delta^-\eta_\Delta^-$, the expression (3) is proved.

□

κ_Δ being symmetric, $a_\Delta^+ = a_\Delta^- = a_\Delta$ and the two summative kernels η_Δ^+ and η_Δ^- can be derived from a unique summative kernel η_Δ: $\forall u \in \Omega,\ \eta_\Delta^+(u) = \eta_\Delta(\frac{\Delta}{2} - u)$ and $\eta_\Delta^-(u) = \eta_\Delta(\frac{\Delta}{2} + u)$. Thus the derivative of the summative kernel $\kappa_\Delta \in \mathcal{K}'(\Omega)$ can be rewritten by:

$$\forall u \in \Omega, -d\kappa_\Delta(u) = a_\Delta \left(\eta_\Delta(\frac{\Delta}{2} - u) - \eta_\Delta(\frac{\Delta}{2} + u)\right), \tag{4}$$

with $a_\Delta = \int_\Omega \max(0, -d\kappa_\Delta(u))du$.

Property 2: Let $a = \int_\Omega \max(0, -d\kappa(u))du$, then, for all $\Delta > 0$:

$$a_\Delta = \frac{a}{\Delta}. \tag{5}$$

Proof First remark that $d\kappa_\Delta(u) = \frac{1}{\Delta^2} d\kappa(\frac{u}{\Delta})$. Thus

$$\begin{aligned} a_\Delta =& \int_\Omega \max(0, -d\kappa_\Delta(u))du = \int_\Omega \max(0, -\frac{1}{\Delta^2}d\kappa(\frac{u}{\Delta}))du, \\ =& \Delta \int_\Omega \max(0, -\frac{1}{\Delta^2}d\kappa(u))du = \frac{1}{\Delta}\int_\Omega \max(0, -d\kappa(u))du, \\ =& \frac{a}{\Delta}. \end{aligned}$$

□

C. Derivative in the sense of distributions

Convolution is a mathematical way of combining two functions to form a third function. The convolution of function s by a summative kernel κ, denoted $\widehat{s}_\kappa$, is given by:

$$\widehat{s}_\kappa(x) = \int_\Omega s(u)\kappa(x-u)du = \int_\Omega s(u)\kappa^x(u)du = \langle s, \kappa^x \rangle, \tag{6}$$

κ^x being the function κ translated in x, and $\langle .,. \rangle$ being the dot product defined for L_1 functions. If the summative kernel κ is derivable, it can be seen as a test function. It is thus possible to link ds, the derivative of s in the sense of distribution, to $d\kappa$, the derivative of κ in the sense of function by:

$$\langle ds, \kappa^x \rangle = \int_\Omega ds(u)\kappa^x(u)du = -\int_\Omega s(u)d\kappa^x(u)du$$
$$= -\langle s, d\kappa^x \rangle .$$

D. From precise to imprecise kernel based expectation

In this section, we introduce summative and maxitive based expectation operators of a function s. An expectation based on a summative kernel κ coincides with the usual expectation based on the probability measure P_κ associated with κ. It is defined by:

$$\mathbb{E}_\kappa(s) = \int_\Omega s dP_\kappa = \int_\Omega s(u)\kappa(u)du. \quad (7)$$

The expectation based on a maxitive kernel π has been introduced in [3]. It uses the conventional extension of the expectation operator, called the Choquet integral [1], and has an interval-valued output. It is defined by:

$$\overline{\underline{\mathbb{E}}}_\pi(s) = \left[\underline{\mathbb{E}}_\pi(s), \overline{\mathbb{E}}_\pi(s)\right], \quad (8)$$

with $\underline{\mathbb{E}}_\pi(s) = \mathbb{C}_{N_\pi}(s)$ and $\overline{\mathbb{E}}_\pi(s) = \mathbb{C}_{\Pi_\pi}(s), \mathbb{C}_{\Pi_\pi}(s)$ (resp. $\mathbb{C}_{N_\pi}(s)$) being the Choquet integral of s with respect to the possibility measure Π_π (resp. the necessity measure N_π). This imprecise valued expectation operator respects the following interesting properties derived from the domination properties defined in section II-A :

$$\forall y \in \overline{\underline{\mathbb{E}}}_\pi(s), \exists \kappa \in \mathcal{M}(\pi) / \mathbb{E}_\kappa(s) = y, \quad (9)$$

and

$$\forall \kappa \in \mathcal{M}(\pi), \mathbb{E}_\kappa(s) \in \overline{\underline{\mathbb{E}}}_\pi(s). \quad (10)$$

This property has been proved in [3].

E. Reformulation of the Parzen Rosenblatt estimator

In this section, we propose a reformulation of the classical Parzen Rosenblatt estimator. This reformulation will be the basis of the new operator we propose.

Let $(x_1, \ldots, x_n)$ be a sample of n i.i.d. observations drawn from a population having an unknown pdf f. The Parzen Rosenblatt kernel estimate [6], [8] of f in every point $x \in \Omega$ is given by:

$$\widehat{f}^n_{\kappa_\Delta}(x) = \frac{1}{n\Delta}\sum_{i=1}^n \kappa\left(\frac{x - x_i}{\Delta}\right) = \frac{1}{n\Delta}\sum_{i=1}^n \kappa^x_\Delta(x_i). \quad (11)$$

Property 3: The estimation $\widehat{f}^n_{\kappa_\Delta}$ in every point $x \in \Omega$ can be rewritten as the dot product kernels κ^x_Δ with the empirical measure e_n:

$$\widehat{f}^n_{\kappa_\Delta}(x) = \langle e_n, \kappa^x_\Delta \rangle, \quad (12)$$

with $e_n = \frac{1}{n}\sum_{i=1}^n \delta^{x_i}$ and δ^{x_i} is the impulse Dirac translated in x_i.

Proof According to (6), we have, for all $x \in \Omega$:

$$\langle e_n, \kappa^x_\Delta \rangle = \int_\Omega \frac{1}{n}\sum_{i=1}^n \delta^{x_i}(u)\kappa^x_\Delta(u)du,$$
$$= \frac{1}{n}\sum_{i=1}^n \int_\Omega \delta^{x_i}(u)\kappa^x_\Delta(u)du,$$
$$= \frac{1}{n}\sum_{i=1}^n \kappa^x_\Delta(x_i) = \widehat{f}^n_{\kappa_\Delta}(x).$$

□

The estimation $\widehat{f}^n_{\kappa_\Delta}$ in every point $x \in \Omega$ can be interpreted as the precise expectation of the empirical distribution e_n according to a neighborhood of x defined by the summative kernel κ_Δ:

$$\widehat{f}^n_{\kappa_\Delta}(x) = \mathbb{E}_{\kappa^x_\Delta}(e_n) = \langle e_n, \kappa^x_\Delta \rangle. \quad (13)$$

Let E_n be the empirical distribution function defined by:

$$\forall x \in \Omega, \quad E_n(x) = \frac{1}{n}\sum_{i=1}^n H(x - x_i), \quad (14)$$

H being the Heaviside function defined by: $H(x) = 1$ if $x \geq 0$ and 0 elsewhere.

e_n being the derivative of E_n in the sense of distributions, the Parzen Rosenblatt estimator can be rewritten, for all $x \in \Omega$, as:

$$\widehat{f}^n_{\kappa_\Delta}(x) = \langle e_n, \kappa^x_\Delta \rangle = \langle dE_n, \kappa^x_\Delta \rangle = \langle E_n, -d\kappa^x_\Delta \rangle. \quad (15)$$

Theorem 1: Let $\kappa_\Delta \in \mathcal{K}'(\Omega)$, whose Jordan decomposition of its first derivative $d\kappa_\Delta$ is: $\forall u \in \Omega, -d\kappa_\Delta(u) = a_\Delta\left(\eta_\Delta(\frac{\Delta}{2} - u) - \eta_\Delta(\frac{\Delta}{2} + u)\right)$, with $a_\Delta \in \mathbb{R}^+$ and $\eta_\Delta \in \mathcal{K}(\Omega)$, then, for all $x \in \Omega$:

$$\widehat{f}^n_{\kappa_\Delta}(x) = a_\Delta\ \mathbb{E}_{\eta_\Delta}(E^{x-}_n - E^{x+}_n),$$

with $\forall u \in \Omega, E^{x-}_n(u) = E_n(x + \frac{\Delta}{2} - u)$ and $E^{x+}_n(u) = E_n(x - \frac{\Delta}{2} + u)$.

Proof According to (15), we have, for all $x \in \mathbb{R}$:

$$\widehat{f}^n_{\kappa_\Delta}(x) = -\mathbb{E}_{d\kappa^x_\Delta}(E_n),$$
$$= \int_{-\infty}^{+\infty} -d\kappa_\Delta(u - x)E_n(u)du,$$
$$= a_\Delta\Big(\int_{-\infty}^{+\infty} \eta_\Delta\Big(\frac{\Delta}{2} - u + x\Big)E_n(u)du$$
$$- \int_{-\infty}^{+\infty} \eta_\Delta\Big(\frac{\Delta}{2} + u - x\Big)E_n(u)du\Big),$$
$$= a_\Delta\Big(\int_{-\infty}^{+\infty} \eta_\Delta(v)E_n\Big(x + \frac{\Delta}{2} - v\Big)dv$$
$$- \int_{-\infty}^{+\infty} \eta_\Delta(v)E_n\Big(x - \frac{\Delta}{2} + v\Big)dv\Big),$$
$$= a_\Delta\left(\int_{-\infty}^{+\infty} \eta_\Delta(v)(E^{x-}_n(v) - E^{x+}_n(v))dv\right),$$
$$= a_\Delta\ \mathbb{E}_{\eta_\Delta}(E^{x-}_n - E^{x+}_n).$$

□

III. Imprecise estimation of density

In this section, we propose an interval-valued estimation of f, the pdf underlying a set of observations $(x_1, \ldots, x_n)$. This new operator is based on defining a particular set of summative kernels and using the know domination properties of the maxitive kernels to derive a very specific interval-valued estimation. This interval-valued estimation contains all the Parzen Rosenblatt estimators obtained with kernels belonging to this set. This set of summative kernel is defined as follows:

$$\mathcal{D}(\pi, a, \Delta) = \left\{ \begin{array}{c} \nu \in \mathcal{K}'(\Omega), \exists \xi \in \mathcal{M}(\pi_\Delta), \text{such that} \\ -d\nu(u) = a_\Delta \left(\xi(\frac{\Delta}{2} - u) - \xi(\frac{\Delta}{2} + u) \right) \end{array} \right\},$$

where $a \in \mathbb{R}^+ (a_\Delta = \frac{a}{\Delta})$, π is a bounded maxitive kernel and $\Delta \in \mathbb{R}^{*+}$ is a bandwidth. Choosing randomly π, a and Δ can lead to an empty set $\mathcal{D}(\pi, a, \Delta)$.
If π, a and Δ are chosen such that there is a summative kernel $\kappa_\Delta \in \mathcal{K}'(\Omega)$ such that: $\forall u \in \Omega, -d\kappa_\Delta(u) = a_\Delta \left(\eta_\Delta(\frac{\Delta}{2} - u) - \eta_\Delta(\frac{\Delta}{2} + u) \right)$, with $\eta_\Delta \in \mathcal{K}(\Omega), a_\Delta \in \mathbb{R}^+$ and for π being a maxitive kernel that dominates η, then, the subset $\mathcal{D}(\pi, a, \Delta)$ is not empty since it contains κ_Δ by construction. Moreover, this set contains kernels whose bandwidth cannot exceed Δ (since the bandwidth of its two derivatives are limited by $\frac{\Delta}{2}$) and cannot be lower that $\frac{\Delta}{2}$. This last configuration corresponds to the uniform kernel of length $\frac{\Delta}{2}$.
Let $\kappa_\Delta \in \mathcal{K}'(\Omega)$ be a summative kernel, whose Jordan decomposition of its first derivative is: $\forall u \in \Omega, -d\kappa_\Delta(u) = a_\Delta \left(\eta_\Delta(\frac{\Delta}{2} - u) - \eta_\Delta(\frac{\Delta}{2} + u) \right)$, with $a_\Delta \in \mathbb{R}^+$ and $\eta_\Delta \in \mathcal{K}(\Omega)$. Let π be the most specific maxitive kernel dominating η [4].

Definition 1: The interval-valued estimation of the pdf, whose empirical distribution is E_n, is defined, for all $x \in \Omega$ by:

$$\overline{\underline{f}}^n_{[\kappa_\Delta]}(x) = a_\Delta \, \overline{\underline{\mathbb{E}}}_{\pi_\Delta}(E_n^{x-} - E_n^{x+}). \quad (16)$$

The specificity of the interval-valued estimation, defined by (16), is due to the following property:

Property 4: Let $\overline{\underline{f}}^n_{[\kappa_\Delta]}$ be the interval-valued estimation, then, for all $x \in \Omega$:

$$\forall \varphi \in \mathcal{D}(\pi, a, \Delta), \widehat{f}^n_\varphi(x) \in \overline{\underline{f}}^n_{[\kappa_\Delta]}(x). \quad (17)$$

Proof According to (10), we have:

$$\forall \xi \in \mathcal{M}(\pi_\Delta), \mathbb{E}_\xi(E_n^{x-} - E_n^{x+}) \in \overline{\underline{\mathbb{E}}}_{\pi_\Delta}(E_n^{x-} - E_n^{x+}),$$

multiplying by a_Δ the expression (17) is proved.
The reverse inclusion defined by:
$\forall y \in \overline{\underline{f}}^n_{[\kappa_\Delta]}(x), \exists\ \xi \in \mathcal{M}(\pi_\Delta)$ and $\varphi \in \mathcal{K}'(\Omega)$ such that $-d\varphi(u) = a_\Delta \left(\xi(\frac{\Delta}{2} - u) - \xi(\frac{\Delta}{2} + u) \right)$ and $y = \widehat{f}^n_\varphi(x)$, is not proved from now.

IV. Computation and Algorithm

This section aims at proposing a practical and efficient computation of the interval-valued estimation proposed in section III. Such a computation is based on the fact that E_n is a step function whose step positions are known. It leads to an exact computation of $\overline{\underline{f}}^n_{[\kappa_\Delta]}(x)$ for any $x \in \Omega$.
Let $\Delta E_n^x = E_n^{x-} - E_n^{x+}$. According to (8), the interval-valued estimation of f, given by (16), can be rewritten by:

$$\begin{aligned} \overline{\underline{f}}^n_{[\kappa_\Delta]}(x) &= a_\Delta \, \overline{\underline{\mathbb{E}}}_{\pi_\Delta}(\Delta E_n^x), \\ &= a_\Delta \left[\mathbb{C}_{N_{\pi_\Delta}}(\Delta E_n^x), \mathbb{C}_{\Pi_{\pi_\Delta}}(\Delta E_n^x) \right], \end{aligned} \quad (18)$$

with $a_\Delta \in \mathbb{R}^+$, $\mathbb{C}_{\Pi_{\pi_\Delta}}(\Delta E_n^x)$ (resp. $\mathbb{C}_{N_{\pi_\Delta}}(\Delta E_n^x)$) being the Choquet integral of ΔE_n^x with respect to the possibility measure Π_{π_Δ} (resp. the necessity measure N_{π_Δ}).
Since E_n is a step function, ΔE_n^x is also step function defined by:

$$\begin{aligned} \forall u \in \Omega,\ \Delta E_n^x(u) =& \frac{1}{n} \sum_{i=1}^n (H(x - x_i + \frac{\Delta}{2} - u) \\ & - H(x - x_i - \frac{\Delta}{2} + u)), \\ =& \frac{1}{n} \sum_{i=1}^n \begin{cases} -1 & \text{if } u \geq x - x_i + \frac{\Delta}{2}, \\ +1 & \text{if } u \leq x_i - x + \frac{\Delta}{2}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

By construction ΔE_n^x has at most $(2n+1)$ different values. Let $\{w_i^x\}_{i \in \{0,\ldots,2n+1\}}$ be the set of $(2n+2)$ value derived from the set of observations, by:

$$w_i^x = \begin{cases} \frac{\Delta}{2} + x_i - x, & \text{if } i \in \{1, \ldots, n\}, \\ \frac{\Delta}{2} + x - x_{i-n}, & \text{if } i \in \{n+1, \ldots, 2n\}, \end{cases} \quad (19)$$

with $w_0^x = e_{min}$ and $w_{2n+1}^x = e_{max}$.
Let $\Theta = \{0, \ldots, 2n\}$ and $(.)$ be the permutation that sorts the w_i^x in ascending order, i.e. $w_{(0)}^x \leq w_{(2)}^x \leq \ldots w_{(2n+1)}^x$. From the set of w_i^x, we can define $(2n+1)$ intervals W_i^x by: $W_i^x = \left\{ [w_{(i)}^x, w_{(i+1)}^x[\right\}_{i \in \Theta}$. By construction, ΔE_n^x is constant on each interval W_i^x.
Let α_n^i be the constant value of ΔE_n^x on W_i^x defined by:

$$\alpha_n^i = \frac{1}{n} \sum_{k=1}^n \begin{cases} -1 & \text{if } c_i \geq x - x_k + \frac{\Delta}{2}, \\ +1 & \text{if } c_i \leq x_k - x + \frac{\Delta}{2}, \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

c_i being the median value of each W_i^x.
Then, ΔE_n^x can be rewritten in:

$$\forall u \in \Omega, \quad \Delta E_n^x(u) = \sum_{i=0}^{2n} \alpha_n^i \mathbb{1}_{W_i^x}(u), \quad (21)$$

$\mathbb{1}_A$ being the indicator function of $A \in \mathcal{P}(\Omega)$ defined by: $\mathbb{1}_A(u) = 1$ if $u \in A$ and 0 elsewhere.
The computation of the interval-valued estimation, defined by (18), involves two Choquet integrals. Due to the step-wise nature of the function to be integrated, the continuous Choquet integral can be computed by a discrete Choquet integral involving a possibility distribution on the $\{W_i^x\}_{i \in \Theta}$.
Let μ_Δ be the discrete possibility distribution induced on Θ

by the continuous possibility distribution π on Ω and defined by:

$$\forall k \in \Theta, \mu_{k\Delta} = \Pi_{\pi_\Delta}(W_k^x). \quad (22)$$

Let $\alpha_n = \left\{\alpha_n^i\right\}_{i\in\Theta}$ be the $(2n+1)$ real values of ΔE_n^x, defined by (20). Let $(.)$ be the permutation that sorts the α_n^i in ascending order, i.e. $\alpha_n^{(0)} \leq \alpha_n^{(1)} \leq \ldots \alpha_n^{(2n)}$. Let $\left\{A_{(i)}\right\}_{i\in\Theta}$ be the $(2n+1)$ subsets of Θ defined by: $A_{(i)} = \{(i), \ldots, (2n)\}$.
The two continuous Choquet integrals of ΔE_n^x with respect to the continuous possibility measure Π_{π_Δ} (resp. necessity measure N_{π_Δ}) on Ω can be computed by the two discrete Choquet integrals of α_n with respect to the discrete possibility measure Π_{μ_Δ} (resp. necessity measure N_{μ_Δ}) on Θ. Such a computation is given by:

$$\mathbb{C}_{\Pi_{\pi_\Delta}}(\Delta E_n^x) = \mathbb{C}_{\Pi_{\mu_\Delta}}(\alpha_n) = \sum_{i=1}^{2n}(\alpha_n^{(i)} - \alpha_n^{(i-1)})\Pi_{\mu_\Delta}(A_{(i)}), \quad (23)$$

and

$$\mathbb{C}_{N_{\pi_\Delta}}(\Delta E_n^x) = \mathbb{C}_{N_{\mu_\Delta}}(\alpha_n) = \sum_{i=1}^{2n}(\alpha_n^{(i)} - \alpha_n^{(i-1)})N_{\mu_\Delta}(A_{(i)}). \quad (24)$$

Computing the interval valued $\overline{\underline{f}}^n_{[\kappa_\Delta]}$ on a location $x \in \Omega$ can be decomposed in 5 steps:

Step 1. Compute the set of $(2n+2)$ values $\{w_i^x\}_{i\in\{0,\ldots,2n+1\}}$ by means of (19).

Step 2. Sort the $\{w_i^x\}$ and compute the $(2n+1)$ intervals W_i^x by: $W_i^x = \left\{[w_{(i)}^x, w_{(i+1)}^x[\right\}_{i\in\{0,\ldots,2n\}}$.

Step 3. Compute the constant values $\left\{\alpha_n^i\right\}_{i\in\{0,\ldots,2n\}}$ by using (20).

Step 4. Compute the two discrete Choquet integrals $\mathbb{C}_{\Pi_{\mu_\Delta}}(\alpha_n)$ and $\mathbb{C}_{N_{\mu_\Delta}}(\alpha_n)$ by means of (23) and (24).

Step 5. Multiply the result obtained in step 4 by a_Δ.

The estimation of the pdf is usually computed on p regularly spaced points of Ω. Let $\{y_j\}_{j\in\{1,\ldots,p\}}$ be those points, the following algorithm uses the procedure described previously to compute the p interval valued estimations of the pdf: $\left\{\left[\underline{f}^n_{[\kappa_\Delta]}(y_j), \overline{f}^n_{[\kappa_\Delta]}(y_j)\right]\right\}_{j\in\{1,\ldots,p\}}$.

Data: the observations $\{x_i\}_{i\in\{1,\ldots,n\}}$, the considered locations $\{y_j\}_{j\in\{1,\ldots,p\}}$, the maxitive kernel π, the bandwidth Δ and the multiplicative factor a_Δ
Result: $\left\{\left[\underline{f}^n_{[\kappa_\Delta]}(y_j), \overline{f}^n_{[\kappa_\Delta]}(y_j)\right]\right\}_{j\in\{1,\ldots,p\}}$
begin
  for $j = 1$ *to* p **do**
  - Compute the set of $(2n+2)$ values $\{w_i^x\}_{i\in\{0,\ldots,2n+1\}}$ (expression (19)).
  - Sort the $\{w_i^x\}_{i\in\{0,\ldots,2n+1\}}$ in increasing order.
  - Compute the $(2n+1)$ intervals $\{W_i^x\}_{i\in\{0,\ldots,2n\}}$.
  - Compute the $\left\{\alpha_n^i\right\}_{i\in\{0,\ldots,2n\}}$ (expression (20)).
  - Sort the $\left\{\alpha_n^i\right\}$ in increasing order and apply the same permutation to the W_i^x.
  - Compute the $\{\mu_{i\Delta}\}_{i\in\{0,\ldots,2n\}}$ (expression (22)).
  - Compute $\mathbb{C}_{\Pi_{\mu_\Delta}}(\alpha_n)$ and $\mathbb{C}_{N_{\mu_\Delta}}(\alpha_n)$:

    $\Pi \leftarrow 0, \ \ \overline{f}^n_{[\kappa_\Delta]}(y_j) \leftarrow 0$
    for $i = 2n$ *to* 1 *down to* 1 **do**
      $\Pi = max(\Pi, \mu_{i\Delta})$
      $\overline{f}^n_{[\kappa_\Delta]}(y_j) = \overline{f}^n_{[\kappa_\Delta]}(y_j) + (\alpha_n^i - \alpha_n^{i-1})\Pi$
    end
    $\overline{f}^n_{[\kappa_\Delta]}(y_j) = a_\Delta \overline{f}^n_{[\kappa_\Delta]}(y_j)$
    $\Pi \leftarrow 0, \ \ \underline{f}^n_{[\kappa_\Delta]}(y_j) \leftarrow 0$
    for $i = 1$ *to* $2n$ **do**
      $\Pi = max(\Pi, \mu_{i\Delta})$
      $\underline{f}^n_{[\kappa_\Delta]}(y_j) = \underline{f}^n_{[\kappa_\Delta]}(y_j) + (\alpha_n^i - \alpha_n^{i-1})\Pi$
    end
    $\underline{f}^n_{[\kappa_\Delta]}(y_j) = a_\Delta \underline{f}^n_{[\kappa_\Delta]}(y_j)$
  end
end

Algorithm 1: Computation of $\left\{\overline{\underline{f}}^n_{[\kappa_\Delta]}(y_j)\right\}_{j\in\{1,\ldots,p\}}$.

V. Experiment

In this experiment we propose to illustrate Property 4 defined in section III. Property 4 says that, if κ_Δ is a symmetric summative kernel with a bounded support and having a first derivative, then, the precise estimate $\hat{f}^n_{\kappa_\Delta}$, defined by (11), is included in the imprecise estimate $\overline{\underline{f}}^n_{[\kappa_\Delta]}$ defined by (16).
In this experiment, the symmetric summative kernel we use is defined on Ω by:

$$\forall x \in \Omega, \ \kappa_\Delta(x) = \frac{1}{2\Delta}(1 + cos(\frac{|x|\,\pi}{\Delta}))\mathbb{1}_{[-\Delta,\Delta]}(x).$$

The summative kernel involved in the Jordan decomposition (4) of the derivative of κ_Δ is given by:

$$\forall x \in \Omega, \ \eta_\Delta(x) = \frac{\pi}{2\Delta}(cos(\frac{|x|\,\pi}{\Delta}))\mathbb{1}_{[-\frac{\Delta}{2},\frac{\Delta}{2}]}(x).$$

The constant value a_Δ involved in this decomposition is equal to $\frac{1}{\Delta}$. We then construct the most specific maxitive kernel dominating η_Δ [4] by:

$$\forall x \in \Omega,\ \pi_\Delta(x) = (1 - sin(\frac{|x|\,\pi}{\Delta}))\mathbb{1}_{[-\frac{\Delta}{2},\frac{\Delta}{2}]}(x).$$

To achieve this experiment, we have drawn 1000 observations $\{x_i\}_{i\in\{1,\ldots,1000\}}$ from a simulated process whose pdf is a mixture of two Gaussian distributions with mean 3 (resp. 8) and variance 1 (resp. 4). The reference interval $\Omega = [-5, 20]$ is divided in 500 equally spaced samples $\{y_i\}_{i\in\{1,\ldots,500\}}$. The value of the bandwidth Δ has been set to 1 since it seems to be adapted to this pdf with this number of observations.

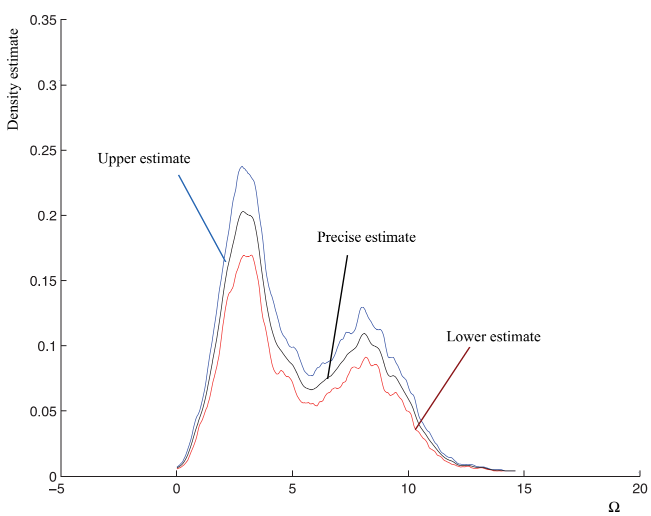

Fig. 1. Superposition, for $\Delta = 1$, of the specific imprecise estimate $\overline{\underline{f}}^n_{[\kappa_\Delta]}$ and the precise estimate $\hat{f}^n_{\kappa_\Delta}$.

Fig. 1 shows the superposition of the precise estimate $\hat{f}^n_{\kappa_\Delta}$ and the specific imprecise estimate $\overline{\underline{f}}^n_{[\kappa_\Delta]}$. The precise estimate is plotted in black, the lower (resp. upper) estimate is plotted in red (resp. blue). As can be seen in Fig. 1, for each value $\{y_i\}_{i\in\{1,\ldots,500\}}$ of the reference interval, $\hat{f}^n_{\kappa_\Delta}(y_i) \in \overline{\underline{f}}^n_{[\kappa_\Delta]}(y_i)$.

VI. Conclusion

In this paper, we have presented a new way for estimating the pdf underlying a finite set of observations with a practical implementation. One of the main characteristic of this estimator is that its output is interval-valued. By construction, this interval-valued probability density estimation is nothing else but the convex set of all the densities that should have been obtained by using the conventional Parzen-Rosenblatt technique with different kernels belonging to a particular set. This set of kernel has a certain interest, since it is the set of all the derivable kernels having a symmetric bounded shape with a bandwidth belonging to a limited interval. Derivable symmetric bounded kernels are among the most used in this context. Thus this set can be instrumental to represent an imprecise knowledge on the shape or bandwidth that has to be used to compute a precise density. Moreover, as shown recently [5] this kind of method can lead to a quantification of the identification error.

A great amount of work remains to answer different questions:

- does this estimator converge and in which sense? For example, how many samples are needed to guarantee a 90% inclusion of the true density in the estimated imprecise density? Does this interval-valued estimate converge to a precise-valued estimate ? From our first attempt it seems that this property is not true,
- does the median of this interval valued pdf have a particular meaning? Does it converge to the true density when the bandwidth tends to 0 and the number of observations tends to infinity?
- can the imprecision of each interval-valued pdf estimation can be considered as a quantification of the estimation error? Can this approach being compared with the traditional method consisting in estimating confidence intervals of a precise-valued pdf estimation [7]?
- is it possible to define a robust distance value between the pdf underlying two sets of observations based on this imprecise estimate?

Future work should also concentrate on practical use of this new density estimate.

References

[1] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, 5:131–295, 1953.
[2] L. de Campos, J. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *international Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.
[3] K. Loquin and O. Strauss. Imprecise functional estimation: the cumulative distribution case. In *SMPS Soft Methods in Probability and Statistics*, pages 175–182, Toulouse, France, Sept. 2008.
[4] K. Loquin and O. Strauss. On the granularity of summative kernels. *Fuzzy sets and systems*, 159:1952–1972, 2008.
[5] K. Loquin and O. Strauss. Noise quantization via possibilistic filtering. In *ISIPTA'09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 297–306, Durham, UK, July 2009.
[6] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065–1076, 1962.
[7] D. Picard and K. Tribouley. Adaptive confidence interval for pointwise curve estimation. *Annals of statistics*, 28:298 – 335, 2000.
[8] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27:832–837, 1956.
[9] L. Schwartz. *Théorie des distributions*. Hermann, Paris, 1950.
[10] D. W. Scott. *Multivariate Density Estimation:Theory, Practice, and Visualization*. Wiley Interscience, New York, 1992.
[11] B. Silvermann. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.
[12] A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer science & business media, New York, 2009.
[13] E. Wegman. Nonparametric probability density estimation. II: A comparison of density estimation methods. *Statistical computation and simulation*, 1:225–245, 1972.