Estimating the variance of a kernel density estimation

Bilal Nehme, Olivier Strauss and Kevin Loquin

Abstract This article proposes an interval-valued extension of kernel density estimation. We show that the imprecision of this interval-valued estimation is highly correlated with the variance of the density estimation induced by the statistical variations of the set of observations.

Keywords Kernel density estimation, maxitive kernel, imprecise expectation.

1 Introduction

The Parzen-Rosenblatt density estimation is a well known nonparametric way of estimating the probability density function (pdf) underlying a finite set of observations. Since the convergence of this estimation towards the true density is only guaranteed for a infinite number of observations, it can be of prime interest to have a measure of the statistical error of this estimation (e.g. its variance). Such a measure cannot be directly computed when the pdf has to be estimated with a single set of observations. One can use resampling techniques, like Jackknife or Bootstrap [4], to perform this estimation. However, those methods can lead to computationally very expensive solutions.

In this paper, we propose a very novel approach for computing this estimation error. This approach is based on an extension of the Parzen-Rosenblatt method that leads to an interval-valued estimation of the pdf. Such an extension have been used in the past [8] for quantifying the effect of the input random noise on the output of a filtering process. It is based on replacing the summative kernel, on which is based the estimation, by a maxitive kernel [7], i.e. a possibility distribution. In this case, however, the Parzen-Rosenblatt estimator has to be reformulated to comply with the maxitive-based estimation extension.

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), 161, rue Ada, 34095 Montpellier Cedex 5 , France. Bilal.Nehme@lirmm.fr, Olivier.Strauss@lirmm.fr.
Télécom ParisTech (Ecole Nationale Supérieure des Télécommunications), 46, rue Barrault, F-75634 Paris Cedex 13, France. kevin.loquin@gmail.com.

2 Preliminarys concepts

This section aims at presenting some preliminaries that are necessary to build the interval-valued pdf estimation we propose. Let Ω be a subset of $\mathbb{R}$, $\mathcal{P}(\Omega)$ the collection of all Lebesgue measurable subsets of Ω and $s : \Omega \to \mathbb{R}$ a bounded L_1 function associated to a distribution in the meaning of Schwartz [12].

We call **summative kernel** [7] a function $\kappa : \Omega \longrightarrow \mathbb{R}^+$ such that $\int_\Omega \kappa(x)dx = 1$. It defines a probability measure on Ω denoted $P_\kappa : \forall A \in \mathcal{P}(\Omega)$, $P_\kappa(A) = \int_A \kappa(x)dx$. Let $\mathcal{K}(\Omega)$ be the set of summative kernels on Ω.

We call **maxitive kernel** [7] a function $\pi : \Omega \longrightarrow [0,1]$ such that $\sup_{x\in\Omega} \pi(x) = 1$. It defines two dual confidence measures on Ω: a possibility measure Π_π and a necessity measure N_π by: $\forall A \in \mathcal{P}(\Omega)$, $\Pi_\pi(A) = \sup_{x\in A} \pi(x)$ and $N_\pi(A) = 1 - \sup_{x\notin A} \pi(x)$. Based on [2], a maxitive kernel π defines a convex set $\mathcal{M}(\pi)$ of summative kernels [6]:

$$\mathcal{M}(\pi) = \{\kappa \in \mathcal{K}(\Omega) / \forall A \in \mathcal{P}(\Omega), N_\pi(A) \le P_\kappa(A) \le \Pi_\pi(A)\}. \quad (1)$$

Let Δ be a positive real value and $x \in \Omega$, a summative kernel κ_Δ^x can be derived from another summative kernel κ by: $\forall u \in \Omega$, $\kappa_\Delta^x(u) = \frac{1}{\Delta}\kappa(\frac{u-x}{\Delta})$. In the same way, a maxitive kernel π_Δ^x can be defined from a maxitive kernel π by: $\forall u \in \Omega$, $\pi_\Delta^x(u) = \pi(\frac{u-x}{\Delta})$. Δ is called the bandwidth of the kernel.

2.1 Derivative of a summative kernel

Kernel used in density estimation are usually unimodal, symmetric with a bounded support and having a first derivative. Let us denote $\mathcal{K}'(\Omega)$ the subset of those kernels on Ω.

Property 1 *Let $\kappa \in \mathcal{K}'(\Omega)$ and $\Delta \in \mathbb{R}^+$, the first derivative $d\kappa_\Delta$ of the kernel κ_Δ can be written as a linear combination of two summative kernels η_Δ^+ and η_Δ^- [9]:*

$$\forall u \in \Omega, -d\kappa_\Delta(u) = a_\Delta \left(\eta_\Delta^+(u) - \eta_\Delta^-(u)\right), \quad (2)$$

where a_Δ is a constant value defined by $a_\Delta = \int_\Omega \max(0, -d\kappa_\Delta(u))$ and $\eta_\Delta^+(u) = \frac{d\kappa_\Delta^+(u)}{a_\Delta}$, $\eta_\Delta^-(u) = \frac{d\kappa_\Delta^-(u)}{a_\Delta}$ with $d\kappa_\Delta^+ = max(0, d\kappa_\Delta)$, $d\kappa_\Delta^- = max(0, -d\kappa_\Delta)$. Note that, by construction, $a_\Delta = \frac{a}{\Delta}$, with $a = \int_\Omega \max(0, -d\kappa(u))du$.

2.2 Derivative in the sense of distributions

The convolution of a L_1 function s by a summative kernel κ, denoted $\widehat{s}_\kappa = s \star \kappa$ is given by [5]:

$$\widehat{s}_\kappa(x) = \big(s \star \kappa\big)(x) = \int_\Omega s(u)\kappa(x-u)du = \int_\Omega s(u)\kappa^x(u)du = \langle s, \kappa^x \rangle \,, \quad (3)$$

κ^x being the function κ translated in x, and $\langle .,. \rangle$ being the dot product defined for L_1 functions. The value $\widehat{s}_\kappa(x)$ can also be viewed as $\mathbb{E}_{\kappa^x}$, the expectation of s according to the neighborhood of x defined by the kernel κ .

If the summative kernel κ is differentiable, it can be seen as a test function [12]. It is thus possible to link ds, the derivative of s in the sense of distributions, to $d\kappa$, the derivative of κ in the sense of functions by [5]:

$$\langle ds, \kappa^x \rangle = \int_\Omega ds(u)\kappa^x(u)du = -\int_\Omega s(u)d\kappa^x(u)du = \langle s, -d\kappa^x \rangle \,. \quad (4)$$

2.3 Reformulation of the Parzen-Rosenblatt density estimator

Let $(x_1, ..., x_n)$ be a sample coming from the same random variable X with density function f. The Parzen-Rosenblatt kernel estimate [10, 11] of the density f in every point $x \in \Omega$ is given by:

$$\widehat{f}^n_{\kappa_\Delta}(x) = \frac{1}{n\Delta}\sum_{i=1}^n \kappa(\frac{x - x_i}{\Delta}) = \frac{1}{n}\sum_{i=1}^n \kappa^x_\Delta(x_i). \quad (5)$$

Property 2 *The estimation $\widehat{f}^n_{\kappa_\Delta}$ in every point $x \in \Omega$ can be interpreted as the expectation of the empirical distribution e_n according to a neighborhood of x defined by the summative kernel κ_Δ:*

$$\widehat{f}^n_{\kappa_\Delta}(x) = \mathbb{E}_{\kappa^x_\Delta}(e_n) = \langle e_n, \kappa^x_\Delta \rangle \,. \quad (6)$$

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. In the same manner, an estimate of the cumulative distribution function F_{η_Δ}, associated with the random variable X, can be obtained by computing the expectation of the empirical distribution function E_n according to a neighborhood of x defined by the summative kernel η_Δ:

$$\widehat{F}^n_{\eta_\Delta}(x) = \mathbb{E}_{\eta^x_\Delta}(E_n) = \langle E_n, \eta^x_\Delta \rangle \,, \quad (7)$$

with $E_n(x) = \frac{1}{n}\sum_{i=1}^{n} H(x - x_i)$ and H being the Heaviside function defined by $H(x) = 1$ if $x \geq 0$ and 0 elsewhere. Since e_n is the derivative of E_n in the sense of distributions [12], the Parzen-Rosenblatt estimator can be rewritten, for all $x \in \Omega$, as:

$$\widehat{f}^n_{\kappa_\Delta}(x) = \langle e_n, \kappa^x_\Delta \rangle = \langle dE_n, \kappa^x_\Delta \rangle = \langle E_n, -d\kappa^x_\Delta \rangle \,. \tag{8}$$

Theorem 1 *Let $\kappa_\Delta \in \mathcal{K}'(\Omega)$, whose first derivative $d\kappa_\Delta$ can be decomposed in: $\forall u \in \Omega, -d\kappa_\Delta(u) = a_\Delta \left(\eta^+_\Delta(u) - \eta^-_\Delta(u)\right)$, with $a_\Delta \in \mathbb{R}^+$ and $\left(\eta^+_\Delta, \eta^-_\Delta\right) \in \mathcal{K}(\Omega)$, then, for all $x \in \Omega$, $\widehat{f}^n_{\kappa_\Delta}(x) = a_\Delta \left(\widehat{F}^n_{\eta^+_\Delta}(x) - \widehat{F}^n_{\eta^-_\Delta}(x)\right)$.*

Proof According to (2) and (7), we have:

$$\widehat{f}^n_{\kappa_\Delta}(x) = a_\Delta \left(\left\langle E_n, \eta^{x+}_\Delta \right\rangle - \left\langle E_n, \eta^{x-}_\Delta \right\rangle\right) = a_\Delta \left(\widehat{F}^n_{\eta^+_\Delta}(x) - \widehat{F}^n_{\eta^-_\Delta}(x)\right). \quad \square$$

3 Interval-valued estimation

A maxitive kernel based imprecise estimate of the cumulative distribution function has been proposed in [6]. It is defined for all $x \in \Omega$ by:

$$\overline{\underline{F}}^n_{\pi_\Delta}(x) = \left[\underline{F}^n_{\pi_\Delta}(x), \overline{F}^n_{\pi_\Delta}(x)\right] = \overline{\underline{\mathbb{E}}}_{\pi^x_\Delta}(E_n) = \left[\underline{\mathbb{E}}_{\pi^x_\Delta}(E_n), \overline{\mathbb{E}}_{\pi^x_\Delta}(E_n)\right], \tag{9}$$

where π is a maxitive kernel, $\Delta \in \mathbb{R}^+$ a bandwidth and $\overline{\underline{\mathbb{E}}}_\pi(.)$ is the imprecise expectation based on the maxitive kernel π [6].

The computation of the lower and the upper bounds of the imprecise cumulative distribution estimator, defined by (9), is given in [6] by:

$$\overline{\mathbb{E}}_{\pi^x_\Delta}(E_n) = \mathbb{C}_{\Pi_{\pi^x_\Delta}}(E_n) = \frac{1}{n}\sum_{i=1}^{n} \left(\pi^x_\Delta(x_i) H(x_i - x) + H(x - x_i)\right), \tag{10}$$

$$\underline{\mathbb{E}}_{\pi^x_\Delta}(E_n) = \mathbb{C}_{N_{\pi^x_\Delta}}(E_n) = \frac{1}{n}\sum_{i=1}^{n} \left((1 - \pi^x_\Delta(x_i)) H(x - x_i)\right), \tag{11}$$

$\mathbb{C}_{\Pi_{\pi^x_\Delta}}(E_n)$ (resp. $\mathbb{C}_{N_{\pi^x_\Delta}}(E_n)$) being the Choquet integral of E_n with respect to the possibility measure $\Pi_{\pi^x_\Delta}$ (resp. the necessity measure $N_{\pi^x_\Delta}$). As shown in [6] when $\kappa \in \mathcal{M}(\pi)$, then $\forall \Delta \in \mathbb{R}^+$, $\forall x \in \Omega$, $\widehat{F}^n_{\kappa_\Delta}(x) \in \overline{\underline{F}}^n_{\pi_\Delta}(x)$.

3.1 Interval-valued estimation of the probability density function

The idea underlying the maxitive based imprecise estimation of the density is the following: instead of dominating the summative kernel on which is based the density estimation like in (9), we will dominate the summative kernels involved in the decomposition (2) of its derivative.
Let $(x_1, \ldots, x_n)$ be a set of n observations, f the pdf underlying the observation process and E_n the empirical distribution function associated with this set of observations. Let $\kappa_\Delta \in \mathcal{K}'(\Omega)$ be a summative kernel, whose derivative $-d\kappa_\Delta$ can be decomposed in: $a_\Delta\left(\eta_\Delta^+ - \eta_\Delta^-\right)$, $a_\Delta \in \mathbb{R}^+$ and $\left(\eta_\Delta^+, \eta_\Delta^-\right) \in \mathcal{K}(\Omega)$. Let π^+ (rsp. π^-) be the most specific maxitive kernel dominating η^+ (rsp. η^-) [7].

Definition 1. A Parzen-Rosenblatt-like imprecise estimator of the pdf underlying a set of observations, whose empirical cumulative is E_n, is defined by:

$$\forall x \in \Omega, \overline{\underline{f}}^n_{(\kappa_\Delta)}(x) = a_\Delta\left(\overline{\underline{\mathbb{E}}}_{\pi_\Delta^{+x}}(E_n) \ominus \overline{\underline{\mathbb{E}}}_{\pi_\Delta^{-x}}(E_n)\right), \tag{12}$$

where $\ominus$ is the Minkowski difference [1].

The question concerns now the properties of the obtained imprecise estimation. We will first denote $\mathcal{D}\left(a, \Delta, (\pi^+, \pi^-)\right)$ a subset of $\mathcal{K}'(\Omega)$ defined by:

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

where a_Δ, Δ, a, π_Δ^+ and π_Δ^- have been previously defined.

The interval-valued estimation, defined by (12),verifies the following property:

Property 3 *Let $\overline{\underline{f}}^n_{(\kappa_\Delta)}$ be the interval-valued estimation of the pdf defined by Equation (12), then:*

$$\forall x \in \Omega, \forall \varphi \in \mathcal{D}\left(a, \Delta, (\pi^+, \pi^-)\right), \widehat{f}^n_\varphi(x) \in \overline{\underline{f}}^n_{(\kappa_\Delta)}(x). \tag{13}$$

Remark 1 *The reverse property of expression (13) is not true, i.e.:*

$$\exists y \in \overline{\underline{f}}^n_{(\kappa_\Delta)}(x), \forall \varphi \in \mathcal{D}\left(a, \Delta, (\pi^+, \pi^-)\right), y \neq \widehat{f}^n_\varphi(x).$$

3.2 Integrated imprecision of the interval-valued estimation

It would have been nice if the imprecision of the interval-valued density estimate we propose had decreased with Δ and $\frac{1}{n}$. Unfortunately, as we prove here, the integral of the imprecision of $\underline{\overline{f}}^n_{(\kappa_\Delta)}$ depends neither on n nor on Δ. To prove this property, we need the following theorem:

Theorem 2 *Let π_Δ be a maxitive kernel. Let $\epsilon^n_{\pi_\Delta}(x) = \overline{F}^n_{\pi_\Delta}(x) - \underline{F}^n_{\pi_\Delta}(x)$ be the imprecision at x of the interval-valued estimation $\underline{\overline{F}}^n_{\pi_\Delta}(x)$, defined by (9), then: $\int_\Omega \epsilon^n_{\pi_\Delta}(x)dx = \rho(\pi_\Delta) = \Delta\ \rho(\pi)$, with $\rho(\pi) = \int_\Omega \pi(x)dx$ being the granulosity of the maxitive kernel π [7], i.e. its degree of imprecision.*

Proof According to (10) and (11), we have $\epsilon^n_{\pi_\Delta}(x) = \frac{1}{n}\sum_{i=1}^n (\pi^x_\Delta(x_i) - \mathbb{1}_{x=x_i})$. Since $\int_\Omega \mathbb{1}_{x=x_i}dx = 0$, $\forall i \in \{1,\ldots,n\}$, we obtain: $\int_\Omega \epsilon^n_{\pi_\Delta}(x)dx = \frac{1}{n}\sum_{i=1}^n \rho(\pi_\Delta) = \Delta\ \rho(\pi)$. $\square$

Theorem 3 *Let $\kappa_\Delta \in \mathcal{K}'(\Omega)$ be a summative kernel. Let $\zeta^n_{(\kappa_\Delta)}(x) = \overline{f}^n_{(\kappa_\Delta)}(x) - \underline{f}^n_{(\kappa_\Delta)}(x)$ be the imprecision at x of the interval-valued estimation $\underline{\overline{f}}^n_{(\kappa_\Delta)}$ defined by (12), then $\int_\Omega \zeta^n_{(\kappa_\Delta)}(x)dx$ is a constant value that we call α.*

Proof According to (12) and by theorem 2 we have:

$$\begin{aligned}\int_\Omega \zeta^n_{(\kappa_\Delta)}(x)dx &= a_\Delta\left(\int_\Omega (\overline{F}^n_{\pi^+_\Delta}(x) - \underline{F}^n_{\pi^+_\Delta}(x))dx + \int_\Omega (\overline{F}^n_{\pi^-_\Delta}(x) - \underline{F}^n_{\pi^-_\Delta}(x))\right)dx,\\ &= a\left(\rho(\pi^+) + \rho(\pi^-)\right) = \alpha.\ \square\end{aligned}$$

The main consequence of theorem 3 is that the defined imprecise estimator cannot converge to the true density, i.e. when $\Delta \to 0$ and $n\Delta \to \infty$, $(\overline{f}^n_{(\kappa_\Delta)} - \underline{f}^n_{(\kappa_\Delta)}) \not\to 0$.

4 Link between imprecision and variance

This section is dedicated to an experiment showing that the imprecision $\zeta^n_{(\kappa_\Delta)}(x)$ of the interval-valued estimate $\underline{\overline{f}}^n_{(\kappa_\Delta)}(x)$ can be used to estimate $var(\widehat{f}^n_{\kappa_\Delta}(x))$, the variance of the Parzen-Rosenblatt estimate of f via the kernel κ_Δ. First, as shown by numerous other works, theoretically $var(\widehat{f}^n_{\kappa_\Delta}(x))$ decreases when n and Δ increases. In fact, as stated in [13]:

$$\forall x \in \Omega, \quad var(\widehat{f}^n_{\kappa_\Delta}(x)) \approx (n\Delta)^{-1} f(x) R(\kappa_\Delta), \tag{14}$$

with $R(\kappa_\Delta) = \int_\Omega \kappa_\Delta(x)^2 dx$. Since the integral of $\zeta^n_{(\kappa_\Delta)}$ depends neither on n nor on Δ, the direct value of $\zeta^n_{(\kappa_\Delta)}(x)$ cannot be used directly to estimate $var(\widehat{f}^n_{\kappa_\Delta}(x))$ but should be multiplied by a factor $\gamma(n,\Delta)$ that depends on both n and Δ. Let us suppose this relation to be linear, i.e.:

$$var(\widehat{f}^n_{\kappa_\Delta}(x)) = \mathbb{E}\big(\gamma(n,\Delta)\ \zeta^n_{(\kappa_\Delta)}(x)\big). \tag{15}$$

Thus, by integrating expression (15), we directly obtain $\gamma(n,\Delta) = \frac{R(\kappa_\Delta)}{\alpha n \Delta}$, with $\alpha = \int_\Omega \zeta^n_{(\kappa_\Delta)}(x)dx$. The experiment we report here aims at testing whether $\big(\gamma(n,\Delta)\ \zeta^n_{(\kappa_\Delta)}(x)\big)$ is correlelated or not with $var(\widehat{f}^n_{\kappa_\Delta}(x))$. It is based on simulating a random process whose underlying pdf is a mixture of two Gaussian distributions of mean 3 (resp. 8) and variance 1 (resp. 4). We use the symmetric summative kernel defined by $\kappa_\Delta(x) = \frac{1}{2\Delta}(1+cos(\frac{|x|\pi}{\Delta}))1\!\!1_{[-\Delta,\Delta]}(x)$. The computation of the different values associated with this kernel are: $\eta^+_\Delta(x) = \eta^-_\Delta(x) = \frac{\pi}{2\Delta}(cos(\frac{|x|\pi}{\Delta}))1\!\!1_{[\frac{-\Delta}{2},\frac{\Delta}{2}]}$, $a = 1$, $\alpha \approx 0.7268$ and $R(\kappa_\Delta) = \frac{3}{4\Delta}$. The value of Δ is fixed to $\Delta = 1$, while the number of observations varies from $n = 1000$ to $n = 10000$. For each values of n, we compute 400 different sets of observation. We then estimate both $var(\widehat{f}^n_{\kappa_\Delta}(x))$ and $\mathbb{E}\big(\gamma(n,\Delta)\ \zeta^n_{(\kappa_\Delta)}(x)\big)$ on 500 equally spaced samples of the reference subset $\Omega = [-5, 20]$.

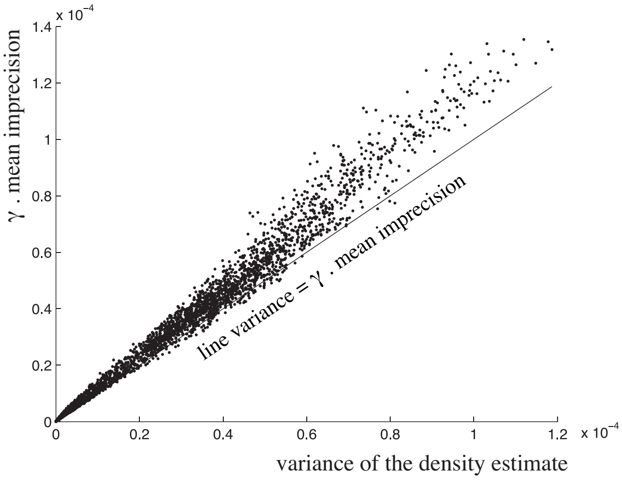

Fig. 1 The cloud of values $\mathbb{E}\big(\gamma(n,\Delta)\ \zeta^n_{(\kappa_\Delta)}\big)$ versus $var(\widehat{f}^n_{\kappa_\Delta})$.

Fig. 1 shows the result of this experiment by plotting $var(\widehat{f}^n_{\kappa_\Delta})$ versus $\mathbb{E}\big(\gamma(n,\Delta)\ \zeta^n_{(\kappa_\Delta)}\big)$. As can be seen on Fig. 1, the correlation between $var(\widehat{f}^n_{\kappa_\Delta})$ and $\mathbb{E}\big(\gamma(n,\Delta)\ \zeta^n_{(\kappa_\Delta)}\big)$ is high (correlation coefficient $r \approx 0.995$). However, the cloud of the computed values is close but rather above the theoretical line materializing equation (15) on Fig. 1. This bias can be explained first by the fact that relation (14) is an approximation and second by the fact that the dependence is possibly not exactly linear. However, the numerous experiments we carried out show that $\big(\gamma(n,\Delta)\ \zeta^n_{(\kappa_\Delta)}\big)$ provide a good estimation of $var(\widehat{f}^n_{\kappa_\Delta})$.

5 Conclusion

The interval-valued nonparametric extension of the kernel density estimation improves on the traditional approach by providing an estimation of the error induced by the statistical variation of the set of observations with a significant increase of the computational complexity.

Future work should focus on the relation between the median of this interval-valued density and the true density (convergence if any ?) and propose a modification of expression (12) that leads to an interval-valued density whose imprecision decreases with the bandwidth of the kernel or when the number of observation increases. We are now working on comparing this approach with the classical approach based on confidence intervals [3].

Acknowledgment

This work is in part supported by VOODDO (2008-2011), a project of ANR and the region of Languedoc Roussillon, France.

References

1. V. Danilov and G. Koshevoy. Cores of cooperative games, superdifferentials of function, and the minkowski difference of sets. *Mathematical Analysis and Applications*, 247:1–14, 2000.
2. D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49(1):65–74, July 1992.
3. E. Gin and R. Nickl. Confidence bands in density estimation. *Annals of Statistics*, 38:1122–1170, 2010.
4. P. Huber. *Robust Statistics*. Wiley, New York, 1981.
5. K. Loquin. *De l'utilisation des noyaux maxitifs en traitement de linformation*. PhD thesis, LIRMM, Université Montpellier II, 2008.
6. 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.
7. K. Loquin and O. Strauss. On the granularity of summative kernels. *Fuzzy sets and systems*, 159:1952–1972, 2008.
8. K. Loquin and O. Strauss. Noise quantization via possibilistic filtering. In *International Symposium on Imprecise Probability: Theory and Applications, ISIPTA 2009*, pages 297–306, Durham, UK, July 2009.
9. B. Nehme and O. Strauss. Towards an interval-valued estimation of the density. In *2010 IEEE International Conference on Computational Intelligence, Accepted paper*, Barcelona, Spain, Sept. 2010.
10. E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065–1076, 1962.
11. M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27:832–837, 1956.
12. L. Schwartz. *Théorie des distributions*. Hermann, Paris, 1950.
13. B. Silvermann. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.