Imprecise functional estimation: the cumulative distribution case

Kevin Loquin, Olivier Strauss

Imprecise functional estimation: the cumulative distribution case

Kevin LOQUIN and Olivier STRAUSS

Abstract In this paper, we propose an adaptation of the Parzen Rosenblatt cumulative distribution function estimator that uses maxitive kernels. The result of this estimator, on every point of the domain of F, the cumulative distribution to be estimated, is interval valued instead of punctual valued. We prove the consistency of our approach with the classical Parzen Rosenblatt estimator, since, according to consistency conditions between the maxitive kernel involved in the imprecise estimator and the summative kernel involved in the precise estimator, our imprecise estimate contains the precise Parzen Rosenblatt estimate.

1 Introduction

The probability density function (pdf) f and the cumulative distribution function (cdf) F of a random variable X on $\Omega \subseteq \mathbb{R}$ are fundamental concepts for describing and representing real data in statistics. These representations are linked by $\forall \omega \in \Omega$, $F(\omega) = \int_{-\infty}^{\omega} f(u)du$. When they cannot be specified, estimates of these functions may be performed by using a sample of n observations independent and identically distributed $(X_1, ..., X_n)$ of X. These observations are summarized by the empirical distribution defined by $e_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$, where δ_{X_i} is the Dirac distribution on X_i or by the empirical cumulative distribution function defined on Ω by $E_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[X_i \leq x]}$, where $\mathbb{1}_A$ is the characteristic function on A.

Different methods have been proposed in the literature for estimating or manipulating the pdf or the cdf underlying a sample of observations. The Parzen Rosenblatt method is one of the most efficient non-parametric techniques [10, 11]. It belongs to the class of functional estimation methods.

Kevin LOQUIN
LIRMM, 161 rue Ada 34392 Montpellier cedex 5, France e-mail: Kevin.Loquin@lirmm.fr

Olivier STRAUSS
LIRMM, 161 rue Ada 34392 Montpellier cedex 5, France e-mail: Olivier.Strauss@lirmm.fr

Generally speaking, functional estimation [1] consists of estimating, for all $x \in \Omega$, a function $h : \Omega \longrightarrow \mathbb{R}$ from another function $g : \Omega \longrightarrow \mathbb{R}$ related to h. The nature of this relation between h and g can take different form: g can be replaced by a sequence $(g_n)_{n>0}$, such that $g_n \longrightarrow h$, when $n \longrightarrow +\infty$, h can be a modification of g (or h is a filtered signal obtained from the signal g), or g can be a discretization of h that has to be recovered by interpolation. So, the estimate of h, at $x \in \Omega$, is function of g and x, which can be expressed as $\hat{h}(x) = \varphi(g,x)$.

For the Parzen Rosenblatt pdf estimator, the function g is the empirical distribution e_n, h is the pdf to be estimated f. The estimator $\hat{h}$ is defined for all $x \in \Omega$, by:

$$\hat{h}(x) = f_{n\kappa_\Delta}(x) = \frac{1}{n}\sum_{i=1}^{n} \kappa_\Delta(x - X_i), \tag{1}$$

with κ the kernel used to perform this estimate and Δ the bandwidth. Note that $\kappa_\Delta(x) = \frac{1}{\Delta}\kappa(\frac{x}{\Delta})$. When g is the empirical cumulative distribution E_n, the cdf F is the function h to be estimated and the estimator $\hat{h}$ is defined for all $x \in \Omega$, by:

$$\hat{h}(x) = F_{n\kappa_\Delta}(x) = \int_{-\infty}^{x} f_{n\kappa_\Delta}(u)du. \tag{2}$$

In the Parzen Rosenblatt like methods, and more generally in all the functional estimation methods, the particular role of the kernel is to define a neighborhood that can be shifted to any location of Ω. The classical (precise) approach makes use of summative kernels. A summative kernel can be seen as a probability distribution, defining a probabilistic neighborhood around each location x of Ω.

This paper considers a new approach (imprecise) that makes use of maxitive kernels. A maxitive kernel can be seen as a possibility distribution, defining a possibilistic neighborhood around each location x of Ω. The main consequence of replacing a summative kernel by a maxitive kernel is that the estimated value is an interval $[\underline{h}(x), \overline{h}(x)]$, instead of a single value $\hat{h}(x)$. We are interested in the relation between the point estimate obtained with the classical approach and the interval estimate obtained with our approach.

The paper is organized as follows. In section 2 we present the classical functional estimation using a summative kernel. In section 3, functional estimation with maxitive kernels is exposed. In section 4, the imprecise functional estimation is presented and mathematically justified. In section 5, we apply our method to the Parzen Rosenblatt cdf estimator. Before concluding, we discuss in section 6 of the choice of the involved maxitive kernel. The method is illustrated by an experiment.

2 Functional estimation with summative neighborhoods

In functional estimation, a summative kernel can be considered as a weighted neighborhood of a given location, called its mode, formally similar to a probability distribution.

Definition 1 *Summative kernels are* $\mathbb{R}^+$*-valued functions* κ *defined on a domain* Ω*, verifying the summativity property:* $\int_\Omega \kappa(x)dx = 1$.

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

$$\kappa_\Delta^u(\omega) = \frac{1}{\Delta}\kappa(\frac{\omega - u}{\Delta}),\ \forall \omega \in \Omega. \tag{3}$$

When seen as a probability distribution, a summative kernel κ has a relevant meaning in the scope of uncertainty theories. It induces a probability measure given by $P_\kappa(A) = \int_A \kappa(\omega)d\omega,\ \forall A \subseteq \Omega$. The value $P_\kappa(A)$ can be interpreted as the degree of probability for a realization of the underlying uncertain phenomenon to fall in A.

Estimation of a given function of $h : \Omega \to \mathbb{R}$ in a summative neighborhood κ_Δ^x of a given location x with bandwidth Δ is given by the expectation of its related function g according to the probability measure $P_{\kappa_\Delta^x}$:

$$\hat{h}(x) = \mathbb{E}_{\kappa_\Delta^x}(g). \tag{4}$$

This approach can be found in [1] for functional estimation in statistics. [7] presents digital signal processing methods that can be reformulated as functional estimators (4).

3 Functional estimation with maxitive neighborhoods

A maxitive kernel is also a weighted neighborhood of a given location, called its mode, formally similar to a possibility distribution or membership function of a normalized fuzzy subset [3].

Definition 2 *A maxitive kernel is a* $[0,1]$*-valued function* π*, defined on a domain* Ω*, verifying the maxitivity property:* $\sup_{\omega\in\Omega} \pi(\omega) = 1$.

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 $\theta = (u, \Delta)$, with u a translation factor and Δ its bandwidth. Any element of this family is obtained, for $u \in \Omega$ and $\Delta > 0$, by

$$\pi_\Delta^u(\omega) = \pi(\frac{\omega - u}{\Delta}),\ \forall \omega \in \Omega. \tag{5}$$

A possibility distribution π has a relevant meaning in the scope of uncertainty theories. π induces a possibility measure given by $\Pi_\pi(A) = \sup_{\omega\in A} \pi(\omega),\ \forall A \subseteq \Omega$. The value $\Pi_\pi(A)$ can be interpreted as the degree of possibility for a realization of the underlying uncertain phenomenon to fall in A.

Now, when the summative neighborhood κ_Δ^x is replaced by a maxitive neighborhood π_Δ^x of a given location x with bandwidth Δ, the Lebesgue integral in estimator (4) has to be replaced by the Choquet integral [2, 9] of g.

4 Imprecise functional estimation

A possibility measure is a special case of concave Choquet capacity ν [15]. The conjugate ν^c of such a capacity, defined by $\nu^c(A) = 1 - \nu(A^c), \forall A \subseteq \Omega$, is a convex capacity. A concave capacity ν can encode a special family of probability measures, noted $core(\nu)$ and defined by

$$core(\nu) = \{P_\kappa, \mid \forall A \subseteq \Omega, \nu^c(A) \leq P_\kappa(A) \leq \nu(A)\}. \tag{6}$$

David Schmeidler and Dieter Denneberg proved the following theorem ([12] proposition 3 and [2] proposition 10.3) for capacities.

Theorem 3. *The capacity ν is concave if and only if for all g such that $\mathbb{C}_\nu(|g|) < +\infty$, then $\forall \kappa | P_\kappa \in core(\nu)$, $\mathbb{C}_{\nu^c}(g) \leq \mathbb{E}_\kappa(g) \leq \mathbb{C}_\nu(g)$.*

From Theorem 3, since a maxitive kernel defines a possibility measure, a maxitive kernel-based estimation of h, generalizing expression (4) is interval valued. The upper and lower bounds are the Choquet integrals of g computed with respectively $\Pi_{\pi_\Delta^x}$ and $N_{\pi_\Delta^x}$, which are capacities (or non additive measures) associated to π_Δ^x a maxitive neighborhood of x, with bandwidth Δ. $N_{\pi_\Delta^x}$ is the conjugate of the possibility measure $\Pi_{\pi_\Delta^x}$, called a necessity measure. These remarks leads to the following corollary of Theorem 3.

Corollary 1. Imprecise functional estimation
Let π be a maxitive kernel, then $\forall x \in \Omega$ and $\forall \Delta > 0$,

$$\forall \kappa | P_\kappa \in core(\Pi_{\pi_\Delta^x}), \ \mathbb{C}_{N_{\pi_\Delta^x}}(g) \leq \mathbb{E}_\kappa(g) \leq \mathbb{C}_{\Pi_{\pi_\Delta^x}}(g). \tag{7}$$

Imprecise estimation of a given function of $h : \Omega \to \mathbb{R}$ in a maxitive neighborhood π_Δ^x of a given location x with bandwidth Δ is given by the Choquet integrals of its related function g according to the possibility and necessity measures $\Pi_{\pi_\Delta^x}$ and $N_{\pi_\Delta^x}$:

$$[\underline{h}(x), \overline{h}(x)] = [\mathbb{C}_{N_{\pi_\Delta^x}}(g), \mathbb{C}_{\Pi_{\pi_\Delta^x}}(g)]. \tag{8}$$

According to Corollary 1, an estimate $\hat{h}(x)$ of h obtained with a summative kernel κ, such that P_κ belongs to $core(\Pi_{\pi_\Delta^x})$, belongs to the estimated interval (8). Besides, the estimation bounds are attained, i.e. there exist two summative kernels η and μ, whose associated probability measures P_η and P_μ are in $core(\Pi_{\pi_\Delta^x})$, such that $\mathbb{E}_\eta(g) = \mathbb{C}_{N_{\pi_\Delta^x}}(g)$ and $\mathbb{E}_\mu(g) = \mathbb{C}_{\Pi_{\pi_\Delta^x}}(g)$.

Replacing a summative kernel by a maxitive kernel for estimating a function h aims at taking into account the imperfect knowledge of the modeler to choose a

particular κ. The specificity [16, 8] of the maxitive kernel chosen by the modeler for performing this imprecise estimation reflects his knowledge. The most specific is the maxitive neighborhood, the smallest is the encoded set. Indeed, if π is more specific than π', some summative kernels encoded by π' will not be encoded by π. The smaller is the encoded set of summative neighborhoods, the closer are the estimation bounds with this method.

5 Imprecise cumulative distribution function estimation

The Parzen Rosenblatt density estimator (1) can be expressed as the estimation of the pdf f, with the empirical distribution $g = e_n$ (summarizing the observations) according to a summative neighborhood κ_Δ^x (see expression (3)):

$$f_{n\kappa_\Delta}(x) = \mathbb{E}_{\kappa_\Delta^x}(e_n). \tag{9}$$

Corollary 1, associated with expression (9) suggests that an imprecise estimation of the Parzen Rosenblatt pdf estimator should be performed by computing the Choquet integral of the empirical distribution e_n according to a maxitive kernel (encoding a family of summative kernels). This direct approach is however not applicable here, since the Choquet integral of the empirical distribution does not exist. Indeed, the computation of this integral only exists for bounded functions. The empirical distribution is not bounded. Actually, the Dirac delta functions, forming e_n, are not functions but mathematical constructions, called distributions.

Nevertheless, the Parzen Rosenblatt cdf estimator (2) involves the empirical cumulative distribution E_n, which is a bounded function. Theorem 4 expresses the Parzen Rosenblatt cdf estimate at a point x, as the estimation of the cdf F with the cumulative empirical distribution $g = E_n$ according to a summative neighborhood of x, κ_Δ^x.

Theorem 4. *Let κ be a summative kernel and $\Delta > 0$ and $n > 0$, then $\forall x \in \Omega$,*

$$F_{n\kappa_\Delta}(x) = \mathbb{E}_{\kappa_\Delta^x}(E_n). \tag{10}$$

Proof. First, note that $f_{n\kappa_\Delta}(x) = \int_\Omega \kappa_\Delta(\omega) e_n(x-\omega)d\omega$. Indeed, $\int_\Omega \kappa_\Delta(\omega)e_n(x-\omega)d\omega = \frac{1}{n}\sum_{i=1}^n \int_\Omega \kappa_\Delta(\omega)\delta_{X_i}(x-\omega)d\omega = \frac{1}{n}\sum_{i=1}^n \kappa_\Delta(x-X_i)$. Thus,

$$\begin{aligned} F_{n\kappa_\Delta}(x) &= \int_{-\infty}^x \Big(\int_\Omega \kappa_\Delta(\omega) e_n(u-\omega)d\omega \Big) du, \\ &= \int_\Omega \Big(\int_{-\infty}^x e_n(u-\omega)du \Big) \kappa_\Delta(\omega)d\omega, \end{aligned}$$

E_n is the cumulative distribution associated to the empirical distribution, i.e. $E_n(\omega) = \int_{-\infty}^\omega e_n(u)du$. Then by successive changes of variable $v := u-\omega$ and $t := x-\omega$, we obtain:

$$\begin{aligned} F_{n\kappa_\Delta}(x) &= \int_\Omega E_n(x-\omega)\kappa_\Delta(\omega)d\omega, \\ &= \int_\Omega E_n(t)\kappa_\Delta(x-t)dt, \\ &= \mathbb{E}_{\kappa_\Delta^x}(E_n). \end{aligned}$$

Since E_n is bounded, an imprecise estimation of F at x can be obtained with a maxitive kernel π_Δ^x.

Theorem 5. *Let π be a maxitive kernel, then $\forall x \in \Omega$, $\forall n > 0$ and $\forall \Delta > 0$,*

$$\forall \kappa_{\Delta'} | P_{\kappa_{\Delta'}} \in core(\Pi_{\pi_\Delta^x}),\ \mathbb{C}_{N_{\pi_\Delta^x}}(E_n) \leq F_{n\kappa_{\Delta'}}(x) \leq \mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n). \tag{11}$$

We now present the computation of the imprecise Parzen Rosenblatt cdf estimate. First, observe that E_n is a simple function that can be expressed on Ω by $E_n(\omega) = \sum_{i=1}^n \frac{i}{n}\mathbb{1}_{[X_{(i)},X_{(i+1)}]}$, where $(.)$ indicates a permutation of the observations such that $X_{(i)} \leq X_{(i+1)}$. Thus, the Choquet integral of E_n can be rewritten as $\mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n) = \frac{1}{n}\sum_{i=1}^n \Pi_{\pi_\Delta^x}(\{\omega \in \Omega : E_n(\omega) \geq \frac{i}{n}\})$. It can easily be observed that $\{\omega \in \Omega : E_n(\omega) \geq \frac{i}{n}\} = \{\omega \in \Omega : \omega \geq X_{(i)}\}$. Since the summation does not depend on the order of the summed elements, $\mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n) = \frac{1}{n}\sum_{i=1}^n \Pi_{\pi_\Delta^x}(\{\omega \in \Omega : \omega \geq X_i\})$. With similar developments on $\mathbb{C}_{N_{\pi_\Delta^x}}(E_n)$, we obtain:

$$\begin{aligned} \mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n) &= \frac{1}{n}\sum_{i=1}^n \Big(1 - N_{\pi_\Delta^x}(\{\omega \in \Omega : \omega < X_i\})\Big), \\ \mathbb{C}_{N_{\pi_\Delta^x}}(E_n) &= \frac{1}{n}\sum_{i=1}^n \Big(1 - \Pi_{\pi_\Delta^x}(\{\omega \in \Omega : \omega < X_i\})\Big). \end{aligned}$$

As exposed in [3, 5], $\underline{F_{\pi_\Delta^x}}(u) = N_{\pi_\Delta^x}(\{\omega \in \Omega : \omega < u\})$ is the lower cdf of the set of cdf associated to the summative kernels of $core(\Pi_{\pi_\Delta^x})$. It is the lower cdf of a p-box [6], whose upper cdf is given by $\overline{F_{\pi_\Delta^x}}(u) = \Pi_{\pi_\Delta^x}(\{\omega \in \Omega : \omega < u\})$. As shown in [5], we have:

$$\underline{F_{\pi_\Delta^x}}(u) = \begin{cases} 0 & \text{if } u < x, \\ 1 - \pi_\Delta^x(u) & \text{otherwise,} \end{cases} \text{ and } \overline{F_{\pi_\Delta^x}}(u) = \begin{cases} \pi_\Delta^x(u) & \text{if } u < x, \\ 1 & \text{otherwise.} \end{cases}$$

We thus obtain the imprecise cdf estimate:

$$\mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n) = \frac{1}{n}\sum_{i=1}^n \Big(\pi_\Delta^x(X_i)\mathbb{1}_{[x \leq X_i]} + \mathbb{1}_{[x > X_i]}\Big), \tag{12}$$

$$\mathbb{C}_{N_{\pi_\Delta^x}}(E_n) = \frac{1}{n}\sum_{i=1}^n \Big((1 - \pi_\Delta^x(X_i))\mathbb{1}_{[x \geq X_i]}\Big). \tag{13}$$

6 Experiment and choice of a maxitive kernel

As in the case of the summative kernel methods, the problem of the choice of a particular maxitive kernel for performing imprecise functional estimation can be discussed. The choice of the summative kernel shape κ is often considered as insignificant in the non-parametric statistics community. The main argument is that the asymptotic behavior (when $n \to +\infty$) of $F_{n\kappa_\Delta}$ and $f_{n\kappa_\Delta}$ depend more on Δ than on the choice of κ [14, 1]. However, the asymptotic conditions are barely fulfilled. In non-asymptotic conditions, the shape of the estimate strongly depend on the shape of κ. Moreover, the knowledge of the modeler is generally insufficient for choosing the appropriate kernel. Instead of choosing one particular summative kernel, we propose to the modeler to choose a family of summative kernels matching his knowledge via the choice of a maxitive kernel.

In such kernel methods, where a summative kernel is considered as a neighborhood, it seems sensible to assume that the chosen basic kernel to be shifted and dilated with expression (3) is centered, even and with a support included in $[-1,1]$. Therefore, it naturally leads to choose a basic maxitive kernel π encoding these particular summative kernels. As shown in [4], the triangular maxitive kernel T is the most specific of such maxitive kernels. The triangular possibility distribution is defined on Ω by $T(\omega) = (1-|\omega|)\mathbb{1}_{[|\omega|\leq 1]}$. We now illustrate Theorem 5 by per-

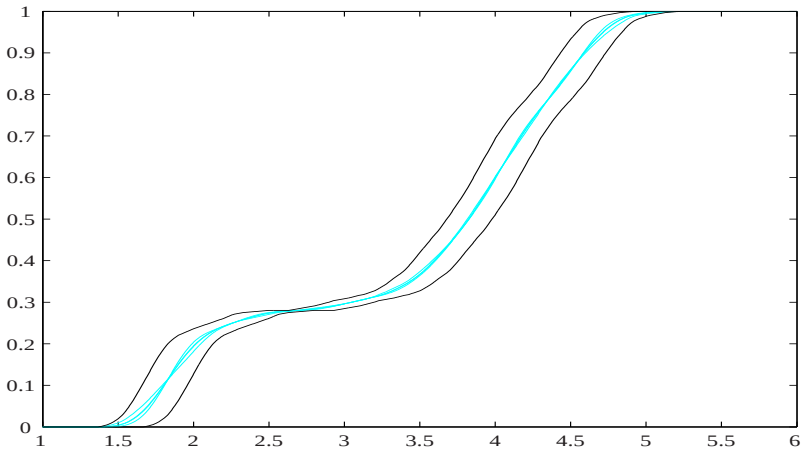

Fig. 1 Imprecise cumulative distribution estimate

forming the summative and maxitive estimates of the cdf underlying a set of 107 observations of the duration in minutes of the eruptions of the Old Faithful geyser in Yellowstone National Park.[1] Each precise estimate has been performed by using four different summative kernels κ_Δ: uniform, Epanechnikov, triweight and cosine

[1] This example, taken from [13], is a popular benchmark in nonparametric estimation.

kernels, with $\Delta = 0.3$. The definitions of the used kernels can be found in [8]. The imprecise estimate is obtained by using a triangular maxitive kernel T with the same Δ. As illustrated on Figure 1, every precise estimates of the cumulative distribution are included in the imprecise estimation interval.

7 Conclusion

In this paper, we proposed an extension of the Parzen Rosenblatt cdf estimate, which takes into account a possible lack of knowledge of the appropriate summative kernel to be involved. Compared to the classical method, our method results in an interval estimate instead of a point estimate. The imprecision of the obtained estimate consistently reflects the lack of knowledge of the modeler, quantified by the specificity of the involved maxitive kernel. We put this sensible imprecise cdf estimation into a wider framework of imprecise functional estimation. Now, the next significant step, in soft statistics, is likely to be the imprecise estimation of the pdf.

References

1. Bosq, D., Lecoutre, J.P.: Théorie de lestimation fonctionnelle. Economica, Paris (1987)
2. Denneberg, D.: Non Additive Measure and Integral. Kluwer Academic Publishers (1994)
3. Dubois, D., Prade, H.: Theorie des possibilités : Applications ‘a la représentation des connaissances en informatique. Masson (1988)
4. Dubois, D., Prade, H., Foulloy, L., Mauris, G.: Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. Reliable Computing **10**, 273-297 (2004)
5. Dubois, D.: Possibility theory and statistical reasonning. Computational Statistics and Data Analysis **51**, 47-69 (2006)
6. Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S., Sentz, K.: Constructing probability boxes and dempster-shafer structures. In: sand2002-4015. Technical report, Sandia National Laboratories, Albuquerque, New Mexico (2002)
7. Jiri Jan: Digital Signal Filtering, Analysis and Restoration. IET (2000)
8. Loquin, K., Strauss, O.: On the granularity of summative kernels. Fuzzy Sets and Systems (2008), doi:10.1016/j.fss.2008.02.021
9. Marichal, J.L.: Aggregation Operations for Multicriteria Decision Aid. In: PhD thesis, Department of Mathematics, University of Liège, Liège, Belgium (1998)
10. Parzen, E.: On estimation of a probability density function and mode. The Annals of Mathematical Statistics **33**, 1065-1076 (1962)
11. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics **27**, 832-837 (1956)
12. Schmeidler, D.: Integral representation without additivity. In Proc. of the Am. Math. Soc. **97**, 255-261 (1986)
13. Silvermann, B. W.: Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability 26. Chapman and Hall (1986)
14. Simonoff, J.S.: Smoothing Methods in Statistics. Springer-Verlag (1996)
15. Walley, P.: Towards a unified theory of imprecise probability. International Journal of Approximate Reasoning **24(2-3)**, 125-148 (2000)
16. Yager, R.R.: Entropy and specificity in a mathematical theory of evidence. International Journal of General Systems **9**, 249-260 (1983)