ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

FUZZY sets and systems

www.elsevier.com/locate/fss

Quasi-continuous histograms

Olivier Strauss*

LIRMM, Université Montpellier II, 161, rue Ada, 34392 Montpellier Cedex 5, France

Received 19 January 2006; received in revised form 5 January 2009; accepted 8 January 2009

Abstract

Histograms are very useful for summarizing statistical information associated with a set of observed data. They are one of the most frequently used density estimators due to their ease of implementation and interpretation. However, histograms suffer from a high sensitivity to the choice of both reference interval and bin width. This paper addresses this difficulty by means of a fuzzy partition. We propose a new density estimator based on transferring the counts associated with each cell of the fuzzy partition to any subset of the reference interval. We introduce three different methods of achieving this transfer. The properties of each method are illustrated with a classic real observation set. The density estimator obtained relates to the Parzen–Rosenblatt kernel density estimation technique. In this paper, we only consider the monovariate case with precise and imprecise observations.

Keywords: Probability density function; Kernels; Fuzzy partition; Possibility theory; Histogram; Granulation

1. Introduction

The histogram is an extremely useful tool for visually summarizing the probability density function (pdf) underlying a data sample. This underlying pdf is the basis of most statistical tools. So it would be natural to include histograms in statistical toolboxes. However, the empirical approximation of the pdf provided by a histogram is quite rough and discontinuous, which can be a major drawback in statistical methods where derivatives of the estimates are required (e.g. mode estimation). In fact, one of the major shortcomings of histogram-based pdf estimation is its sensitivity to the choice of both reference interval and cell number (or bin width).

Over the last decade, several authors have proposed to reduce this sensitivity by replacing the crisp partition by a fuzzy partition. Convergence of such fuzzy-partition-based histograms (fpbh) to the underlying pdf has been proved in mean squared error [21] and integrated mean square error [47]. The reduced sensitivity to the choice of partition compared to the crisp case was shown in [20]. As pointed out by Chielewski in [4], there is an obvious analogy between the weighted voting process performed by an fpbh and the simple low-pass filtering of a conventional crisp-partition-based histogram (cpbh). This low-pass filtering is the most relevant explanation for this reduced sensitivity to the arbitrariness of the choice of partition.

Due to this "good" property, some authors have proposed fuzzy histograms as a statistical computation tool. In [32], Runkler proposes the use of an fpbh to compute a chi-squared test for independence between two data samples.

* Tel.: +33 467 418 587; fax: +33 467 418 500.
E-mail address: Olivier.Strauss@lirmm.fr (O. Strauss).

0165-0114/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2009.01.013

Strauss and Comby use an fpbh in the computation of the centiles [41] and modes [5] of the pdf underlying a data sample. A novel approach to computing conditional probabilities is presented by van den Berg et al. in [44], leading to a firm theoretical foundation for the competitive exception learning algorithm introduced in [43]. Different applications have been proposed for enhancement of classical statistical procedures in image processing, e.g. [2], for image thresholding, and [25], for representing the relative position of objects. One of the main applications of fpbh in image processing is the so-called fuzzy Hough transform (fht), which enables the robust detection of several classes of shapes [3,4,15,40,42].

As noted by Runkler, and most other authors, the fpbh approach outperforms its crisp relatives in terms of smoothness, low complexity, robustness against outliers and low sensitivity to the position of data clusters and to the arbitrariness of the choice of partition. However, these sensitivities still exist and the search for a putative optimal bin width and optimal reference interval position is still relevant. Moreover, most of the fpbh methods proposed in the recent literature do not account for a known defects in the datasets, such as knowledge about their imprecision. Note that the fpbh is usually referred to as a fuzzy histogram in the literature, which is inaccurate since only the partition is fuzzy while the accumulator is crisp. The only attempt to define a histogram whose accumulators are fuzzy has been proposed by Viertl et al. [45]. However, this approach involves fuzzy observations and a crisp partition.

Most statistical tools using fpbh rely on a method that derives a continuous estimate of the underlying pdf from the discrete estimate performed by the histogram. This method generally consists of convolving kernels of the partition with another kernel centered on the value on which the underlying pdf has to be estimated. This method implicitly assumes that accumulations within different bins are independent. This assumption is true when considering a crisp partition, but false when considering a fuzzy partition, since the accumulators associated with two contiguous cells are obviously dependent. This former approach is highly linked with the binned kernel density estimator proposed by Scott and Sheather [35] that mostly relates to kernel density estimation.

In a recent article, Nuñez-Garcia et al. introduced the idea that the construction of a histogram implicitly involves the generation of a sample of random sets [26]. They propose to solve some inherent flaws in the histogram method, such as the recurrent problem of the choice of partition resolution, by using a wide range of techniques related to random sets, including possibility theory and fuzzy measure theory.

In this paper, we propose to solve this recurrent problem in a completely different manner. Rather than trying to find a method for defining an optimal partition, we propose to use the ability of a histogram based on a fuzzy partition of $\mathbb{R}$ to provide an approximate representation of the imprecise probability density underlying the observations. We then extend the usual pignistic transfer to propagate the ambiguity induced by the overlapping of contiguous fuzzy cells, in the weighted voting process, to the estimation of the votes purporting any measurable set of $\mathbb{R}$ covered by the partition. In the same way as Perfilieva [29], we come up with a method that leads to imprecise estimations instead of precise estimations.

We call this association of fpbh with counting transfer techniques *quasi-continuous histograms* (QCH) instead of *fuzzy histograms* because they aim at handling a continuous representation of underlying pdf by discrete operations. QCH were first introduced for performing a mode estimation in a computer vision context [5]. This mode estimation was based on certain conjectural properties concerning accumulation and counting transfer. In this paper, we propose to systematize some of the notions presented in the former paper and to give certain mathematical proofs of the properties induced by this technique. Moreover, we consider the accumulation process not only with precise but also with imprecise observed data.

The structure of this paper is as follows. Section 2 introduces the summative and non-summative kernel concepts, and their link with density estimation. In Section 3, the QCH is introduced as a simple extension of the binary histogram to a fuzzy partition. Section 4 presents the counting transfer as a key feature of the QCH technique. Section 5 is dedicated to illustrative experiments based on the very classic Old Faithful Geyser data [36].

2. Kernels and density estimation

Kernel functions, also simply called kernels, are essential tools in the context of density estimation. They can be seen as weighted neighborhoods ensuring smooth interplay between the continuous domain of the density to be estimated and the discrete domain of the observations. We present, in this section, two kinds of kernel and the way they relate to kernel density estimation.

2.1. Summative and maxitive kernels

Defining a kernel is a way of generalizing the concept of a crisp neighborhood. A kernel induces a preference order for all possible crisp neighborhoods under consideration. A kernel is often considered as a weighted neighborhood, spreading or collecting information around a given location, called its mode. The mode can be a singleton or a set. The extent of this neighborhood is bounded by its support. For most applications in which they are involved, kernels are chosen to be unimodal and symmetric.

A summative kernel [19] is an $\mathbb{R}^+$-valued Lebesgue measurable function κ, defined on an interval $\Omega \subseteq \mathbb{R}$, satisfying the summative normalization condition:

$$\int_{\Omega} \kappa(x)\,\mathrm{d}x = 1, \tag{1}$$

where $\int \kappa(x)\,\mathrm{d}x$ designates the Lebesgue integral of κ over Ω.

The mode of this summative kernel is the set of values at which κ attains its maximum value. Its support is the set of strictly positive values. A unimodal kernel is a kernel whose mode is reduced to a single element.

A summative kernel can be seen as a probability distribution inducing a probability measure, denoted P_κ, and computed in this way:

$$\forall A \subseteq \Omega, \quad P_\kappa(A) = \int_A \kappa(x)\,\mathrm{d}x, \tag{2}$$

where A is a Lebesgue measurable subset of Ω. $P_\kappa(A)$ verifies the Kolmogorov additivity axiom. It can be interpreted as the degree of probability that the occurrence of an event in the neighborhood defined by κ will fall in A.

A maxitive kernel is another way of defining a weighted neighborhood. A maxitive kernel [19] is a normalized fuzzy subset E, defined on a domain Ω, satisfying the maxitive normalization condition:

$$\operatorname*{Sup}_{x\in\Omega} \mu_E(x) = 1. \tag{3}$$

The mode of this maxitive kernel is the set of $x \in \Omega$ values such that $\mu_E(x) = 1$. Its support is the set of values for which $\mu_E(x)$ is strictly positive. A unimodal maxitive kernel is also called a fuzzy number [10].

The membership function of the maxitive kernel E can be seen as a possibility distribution π_E:

$$\forall x \in \Omega, \quad \pi_E(x) = \mu_E(x), \tag{4}$$

inducing a possibility measure [13], denoted Π_E, and computed in this way:

$$\forall A \subseteq \Omega, \quad \Pi_E(A) = \operatorname*{Sup}_{x\in A} \mu_E(x). \tag{5}$$

The value $\Pi_E(A)$ can be interpreted as the degree of possibility that the occurrence of an event in the neighborhood defined by E will fall in A. A possibility measure does not verify the Kolmogorov additivity axiom. Especially for $\Pi_E(A^c) \neq 1 - \Pi_E(A)$, where A^c is the complementary set of A in Ω. Therefore the normalized fuzzy subset E also defines a dual confidence measure, called a necessity measure, denoted N_E, and computed in this way:

$$\forall A \subseteq \Omega, \quad N_E(A) = 1 - \Pi_E(A^c). \tag{6}$$

When a kernel is involved in an information extraction process, its non-specificity highly influences the perceived resolution of the phenomenon considered. For a maxitive bounded kernel E, this non-specificity can be simply and naturally measured by means of the continuous generalization of the cardinality of a fuzzy subset [19]. Due to the relation between this non-specificity and the granularity of the perception of the phenomenon under consideration, we improperly call $\Gamma(E)$ the *granularity* of the maxitive kernel E. It is defined by

$$\Gamma(E) = \int_{\Omega} \mu_E(x)\,\mathrm{d}x. \tag{7}$$

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

A summative kernel κ is said to be dominated by a maxitive kernel E if the probability measure P_κ is dominated by the possibility measure Π_E [6], i.e.

$$\forall A \subseteq \Omega, \quad P_\kappa(A) \leqslant \Pi_E(A). \tag{8}$$

Note that $P_\kappa(A) \leqslant \Pi_E(A)$ entails $N_E(A) \leqslant P_\kappa(A)$.

Thus, a maxitive kernel E can encode a convex family of summative kernels denoted $\Im(E)$ and defined by

$$\forall A \subseteq \Omega, \quad \Im(E) = \{\kappa | \forall A \subseteq \Omega, P_\kappa(A) \leqslant \Pi_E(A)\}. \tag{9}$$

Some transformations are available to associate a maxitive kernel with a summative one and vice versa. The first simple idea consists of dividing the membership function of a maxitive kernel by its granularity to obtain a summative kernel or to divide the probability function associated with the summative kernel by its maximum value to obtain a maxitive kernel. However, this construction is rather arbitrary: the obtained summative kernel may not be dominated by the original maxitive kernel and vice versa. As proposed in [19], the most relevant transformation consists of associating, with a summative kernel, the most specific maxitive kernel that dominates it. This possibility/probability transformation has been proposed by Dubois and Prade [11] and will now be referred to as the Dubois–Prade transformation.

In this context, the triangular symmetric maxitive kernel plays a peculiar role. In fact, as proved by Dubois et al. in [12], the triangular symmetrical fuzzy number whose mode is m and whose spread is Δ is the lowest upper bound of the possibility distribution associated with any symmetric probability distribution of mode m and support $[m - \Delta, m + \Delta]$. Note that in a recent paper [18] Jaulin proposes a very similar approach to compute the inner and outer approximations of minimal-volume credible sets, i.e. the set of minimal volume that contains a desired parameter with a given probability level.

For a more complete introduction to summative and maxitive kernels, see [19].

2.2. Kernel-based density estimation

In statistics, kernel-based density estimation (or the Parzen window method, named after Emanuel Parzen [27]) is a way of estimating the pdf underlying a finite set of observations of a random variable at any point of a considered domain Ω. It has received considerable attention in the literature [16,36,37]. Kernel density estimation can be seen as an attempt to construct a histogram where every point of Ω is the center of a sampling interval, so that the histogram no longer requires a particular choice of cell position. A first approach involves placing a box of width Δ around each point whose density has to be estimated. Let I_x be the interval $[x - \Delta/2, x + \Delta/2)$, let $(x_i)_{i=1,\dots,n}$ be a finite number of n observations, then the naive kernel estimate of $f(x)$, i.e. the underlying pdf to be estimated, is defined by

$$\hat{f}(x) = \frac{1}{n\Delta} \sum_{i=1}^{b} \chi_{I_x}(x_i), \tag{10}$$

where χ_{I_x} is the characteristic function of interval I_x.

The estimated density is still not fully satisfactory in the sense that it is a discontinuous function. The Parzen–Rosenblatt density estimation consists of generalizing this first approach to overcome its step-wise nature by replacing interval I_x by a summative kernel.

The density estimate based on the kernel κ is then defined by

$$\hat{f}_\kappa(x) = \frac{1}{n} \sum_{i=1}^{n} \kappa(x - x_i). \tag{11}$$

Usually, though not always, κ is chosen in the class of symmetrical positive density functions. The smoothness of the estimation is controlled by both the shape of κ and its bandwidth Δ. $\hat{f}_\kappa(X)$, the estimation of $f(x)$, will inherit all the continuity and differentiability properties of κ. In particular, since κ is usually a smooth differentiable function, $\hat{f}_\kappa(x)$ is a smooth curve having derivatives.

Remark 1. Within expression (10), the interval I_x is a maxitive kernel. Therefore another generalization of this expression can be obtained by considering I_x as a fuzzy interval and defining Δ as being the granularity of I_x. This generalization leads to defining

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

$$\hat{f}_\kappa(x) = \frac{1}{n\Delta}\sum_{i=1}^{n}\mu_{I_x}(x_i), \tag{12}$$

where μ_{I_x} is the membership function of interval I_x. Expressions (11) and (12) are equivalent when considering $\kappa(x - x_i) = \mu_{I_x}(x_i)/\Delta$.

Kernel density estimation is one of the most commonly used estimators. Although it has been extensively studied from the mathematical and practical viewpoints, it still has some minor unsolved drawbacks:

- the choice of kernel and its bandwidth is highly arbitrary, but no method has been proposed to provide a measure of the consequences of this choice on the estimation of $f(x)$,
- the local convergence of the estimator is linked with local data density. Nearest neighbor methods [31] propose the adaptation of window size to account for this local density. Since the number of observations is usually low compared to the convergence condition, an error systematically occurs in the estimation of $f(x)$. To date, it seems that no method has been proposed for the measurement, in the estimation of $f(x)$, of error due to the inappropriateness of the kernel with respect to the local observation density,
- a known imprecision in the observation cannot be easily exploited in usual density estimators,
- the practical computation of the density by using the kernel method requires estimation of the density on a fixed partition of the reference interval Ω, but the imprecision due to sampling is usually ignored.

Generally speaking, kernel estimates do not explicitly reflect known or unknown flaws of the observation sample, nor the influence of the arbitrary choice of kernel, kernel bandwidth or reference interval sampling.

3. Quasi-continuous histograms (QCH)

The theoretical framework of QCH is based on fuzzy rough set theory [9]. The main goal is to provide a framework that generalizes the histogram concept by dissociating histogram granularity from the precision of the information computed using this histogram [32,41,44]. QCH are built on a fuzzy partition to reduce the effect of arbitrary partitioning [24,29]. The QCH principle relates to kernel density estimation [27,28,36,37], except that maxitive kernels are considered. The basic ideas of this approach were first proposed in [5]. We restrict our study to the univariate case.

3.1. Accumulation of precise data in a fpbh

Let $(C_k)_{k=1,\dots,p}$ be a uniform partition of p fuzzy cells of the reference interval $\Omega = [x_{\min}, x_{\max}]$ (Fig. 1). The mode of each cell C_k is m_k. The granularity of each cell is Δ, since we consider a regular partition. Δ is then called the granularity of the partition. The distance between two consecutive modes of the partition equals Δ. By convention, we extend the domain Ω by adding two fake modes $m_0 = x_{\min} - \Delta$ and $m_{p+1} = x_{\max} + \Delta$. The fuzzy partition is strong [33] if it fulfills the following conditions:

(1)

$$\forall x \in \Omega, \quad \sum_{k=1}^{p}\mu_{C_k}(x) = 1 \quad \text{(strength condition)}, \tag{13}$$

(2) $\mu_{C_k}(m_k) = 1$ (unimodality),
(3) $\mu_{C_k}(x)$ monotonically increases on $[m_{k-1}, m_k]$ and monotonically decreases on $[m_k, m_{k+1}]$,
(4) $\forall x \in \Omega, \exists! k \in \{1, \dots, p\}$ such that $\mu_{C_k}(x) \neq 0$ and $\mu_{C_{k+1}}(x) \neq 0$,
(5) $\forall x \in [0, \Delta]$, $\mu_{C_k}(m_k + x) = \mu_{C_k}(m_k - x)$ (symmetry).

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

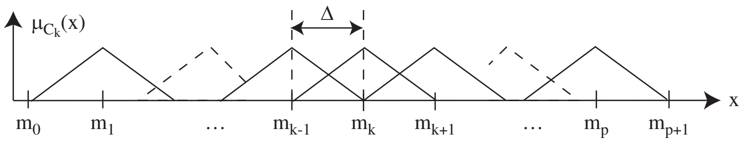


Fig. 1. Fuzzy partition of $\mathbb{R}$.

Let $(x_i)_{i=1,\ldots,n}$ be a finite set of n observations and let $\mu_{C_k}(x)$ be the membership function of C_k, then the value of the accumulator a_k associated with each cell C_k is given by

$$a_k = \sum_{i=1}^{n} \mu_{C_k}(x_i). \tag{14}$$

Each cell C_k can be considered as a maxitive kernel whose membership function is obtained by translating a generic membership function μ_E to the mode of a fuzzy cell:

$$\forall x \in \Omega, \quad \mu_{C_k}(x_i) = \mu_E(x_i - m_k). \tag{15}$$

Remark 2. The accumulators associated with the histogram provide a rough representation of the dataset that induces ambiguities in data position: the information about the real position of each observation x_i is lost and replaced by the imprecise information represented by the cell it belongs to. In fact, each observation votes for two contiguous cells at most. The ambiguity in position is aggravated compared to the crisp case.

Due to the strength condition, the granularity of each cell of the partition is equal to the granularity of the histogram. Note that the triangular membership function is a good candidate for defining the partition due to the peculiar role of triangular numbers in connection with dominated probability functions (see Section 2.1). Moreover, the triangular partition is the only fuzzy partition based on fuzzy numbers fulfilling conditions (13), whose basic membership function is a non-summative kernel associated with a summative kernel (uniform density) by the Dubois–Prade transformation [13] and whose granularity equals the distance between two modes.

Property 1. $\sum_{k=1}^{p} a_k = n.$

Proof. This property is a direct consequence of the strength condition (13).

$$\sum_{k=1}^{p} a_k = \sum_{k=1}^{p}\left(\sum_{i=1}^{n} \mu_{C_k}(x_i)\right) = \sum_{i=1}^{n}\left(\sum_{k=1}^{p} \mu_{C_k}(x_i)\right) = \sum_{i=1}^{n} 1 = n. \qquad \square$$

3.2. Accumulation of imprecise data in a fpbh

In many real life situations, observed data cannot be considered as precise. This kind of situation occurs when the considered data are measurements of continuous quantities, or digitalized (i.e. quantized) observations. An imprecise observation is usually characterized by a domain that is said to contain the true value of the measure with a prescribed confidence level. When the measurement process is inaccurate but is not subject to spurious random variations, then this level equals 1. This is a typical assumption leading to interval-analysis-based guaranteed estimation [17,18]. The consistency of building a cpbh with imprecise data has been studied by Derzko [1] for data whose imprecision is due to quantization. One of the main results is that this kind of histogram does not converge to the underlying pdf but to a smooth estimate of this density involving a sampling function and the quantization step. In [45] Viertl and Trutschnig consider building a histogram with fuzzy observations; the consistency of extending the concept of relative frequencies to the case of fuzzy data is proved in [39]. In this article, we consider building an fpbh with interval-valued observations.

Let $(X_i)_{i=1,\ldots,n}$ be a finite set of n purely imprecise observations. A purely imprecise observation is a domain X_i (usually an interval) whose characteristic function is χ_{X_i}. It can be considered as a random set according to Dempster [7,23]. The value of a_k, the accumulator associated with C_k, cannot be computed using expression (14): the concept of counting the number of x_i belonging to C_k should be replaced by the concept of counting the number of X_i compatible with C_k.

The upper and lower bounds of this compatibility are called, respectively, the possibility and the necessity of X_i restricted to C_k [10]. Those bounds are computed in this way:

$$\Pi_i(C_k) = \operatorname*{Sup}_{x\in\Omega}\{\min(\mu_{C_k}(x), \chi_{X_i}(x))\} = \operatorname*{Sup}_{x\in X_i}\{\mu_{C_k}(x)\} \quad \text{(possibility)}, \tag{16}$$

$$N_i(C_k) = \operatorname*{Inf}_{x\in\Omega}\{\max(\mu_{C_k}(x), 1-\chi_{X_i}(x))\} = \operatorname*{Inf}_{x\notin X_i}\{\mu_{C_k}(x)\} \quad \text{(necessity)}. \tag{17}$$

Consequently, the data imprecision leads to an imprecise counting value. The imprecise accumulation is valued by the interval: $A_k = [\underline{a_k}, \overline{a_k}]$:

$$\underline{a_k} = \sum_{k=1}^{n} N_i(C_k), \quad \overline{a_k} = \sum_{k=1}^{n} \Pi_i(C_k). \tag{18}$$

Remark 3. Expression (18) is a bipolar generalization of expression (14), i.e. if the observations are precise then the accumulation is precise and hence $\underline{a_k} = \overline{a_k}$.

Property 2. $\displaystyle\sum_{i=1}^{p} \underline{a_k} \leqslant n \leqslant \sum_{i=1}^{p} \overline{a_k}.$

Proof. Let $(x_i)_{i=1,\ldots,n}$ be a set of precise values belonging to the imprecise observations, i.e. $\forall i$, $\chi_{X_i}(x_i) = 1$. By definition $\Pi_i(C_k) \geqslant \min(\mu_{C_k}(x_i), \chi_{X_i}(x_i)) = \mu_{C_k}(x_i)$ and $N_i(C_k) \leqslant \max(\mu_{C_k}(x_i), 1-\chi_{X_i}(x_i)) = \mu_{C_k}(x_i)$. Thus,

$$\sum_{i=1}^{p} \overline{a_k} = \sum_{k=1}^{p}\sum_{i=1}^{n} \Pi_i(C_k) = \sum_{i=1}^{n}\sum_{k=1}^{p} \Pi_i(C_k) \geqslant \sum_{i=1}^{n}\sum_{k=1}^{p} \mu_{C_k}(x_i) = n \quad \text{and}$$

$$\sum_{i=1}^{p} \underline{a_k} = \sum_{k=1}^{p}\sum_{i=1}^{n} N_i(C_k) = \sum_{i=1}^{n}\sum_{k=1}^{p} N_i(C_k) \leqslant \sum_{i=1}^{n}\sum_{k=1}^{p} \mu_{C_k}(x_i) = n.$$

Note that this property is also valid for the accumulation of imprecise data in crisp histograms. □

Remark 4. If the granularity of the histogram is lower than twice the granularity of X_i then $\forall k \in \{1, \ldots, p\}$, $N_i(C_k) = 0$. Therefore the granularity of the histogram has to be adapted to the granularity of the data to ensure $\underline{a_k} \neq 0$.

3.3. Imprecise counting and imprecise probabilities

When associated with a counting transfer method (see Section 4), an fpbh leads to an imprecise estimate of the number of votes purporting any interval W of the considered reference interval Ω. Thus, the probability of W may be thought as imprecise [46].

By conventional statistics, the probability of an event can be estimated by its relative frequency, i.e. the ratio of observations in favor of this event to the total number of considered observations. Let n be the total number of observations and $\widehat{\mathrm{nb}}(W)$ the number of observations purporting $W \subseteq \Omega$:

$$\widehat{P}(W) = \frac{\widehat{\mathrm{nb}}(W)}{n}. \tag{19}$$

When the number of observations purporting W is imprecise, $\widehat{\mathrm{nb}}(W)$ has to be replaced by the interval $\widehat{\mathrm{Nb}}(W) = [\underline{\mathrm{nb}}(W), \overline{\mathrm{nb}}(W)]$.

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

In this case, the ratio (19) is imprecise and its upper and lower bounds are [46]

- the upper probability measure of W:

$$\overline{P}(W) = \frac{\overline{\mathrm{nb}}(W)}{\overline{\mathrm{nb}}(W) + \underline{\mathrm{nb}}(W^{c})}, \tag{20}$$

- the lower probability measure of W:

$$\underline{P}(W) = \frac{\underline{\mathrm{nb}}(W)}{\underline{\mathrm{nb}}(W) + \overline{\mathrm{nb}}(W^{c})}. \tag{21}$$

By construction, this imprecise probability measure follows the fundamental property:

$$\overline{P}(W) = 1 - \underline{P}(W^{c}). \tag{22}$$

When $\underline{\mathrm{nb}}(W) = \sum_{i=1}^{n} N_i(W)$ and $\overline{\mathrm{nb}}(W) = \sum_{i=1}^{n} \Pi_i(W)$, then $\underline{\mathrm{nb}}(W) + \overline{\mathrm{nb}}(W^{c}) = n$ and expressions (20) and (21) coincide with the conventional definition of upper and lower relative frequency (see [39] and references therein). When the interval $\widehat{\mathrm{Nb}}(W)$ is just an interval that is said to contain the true value of $\widehat{\mathrm{nb}}(W)$ then expressions (20) and (21) have to be used to compute the interval of all the values of the ratio (19).

Property 3. Let $(X_i)_{i=1,\ldots,n}$ be a finite set of n precise or imprecise observations. Let $\widehat{\mathrm{nb}}(W)$ be a precise estimate and $\widehat{\mathrm{Nb}}(W)$ an imprecise estimate of the number of observations purporting any interval W of the considered reference interval Ω such that $\widehat{\mathrm{nb}}(W) \in \widehat{\mathrm{Nb}}(W)$. Let $\widehat{P}(W)$ be the precise estimation of the probability of W computed with $\widehat{\mathrm{nb}}(W)$ and expression (19). Let $\overline{P}(W)$ (rsp. $\underline{P}(W)$) be the upper (rsp. lower) estimate of the probability of W computed with $\widehat{\mathrm{Nb}}(W)$ and expression (20) (rsp. expression (21)), then $\widehat{P}(W) \in [\underline{P}(W), \overline{P}(W)]$.

Proof. First, let us remark that $\forall W \subseteq \Omega$, $\widehat{\mathrm{nb}}(W) + \widehat{\mathrm{nb}}(W^{c}) = n$, thus

$$\frac{1}{\widehat{P}(W)} = \frac{\widehat{\mathrm{nb}}(W) + \widehat{\mathrm{nb}}(W^{c})}{\widehat{\mathrm{nb}}(W)} = 1 + \frac{\widehat{\mathrm{nb}}(W^{c})}{\widehat{\mathrm{nb}}(W)}.$$

Since $\widehat{\mathrm{nb}}(W) \geqslant \underline{\mathrm{nb}}(W)$ and $\widehat{\mathrm{nb}}(W^{c}) \leqslant \overline{\mathrm{nb}}(W^{c})$,

$$\frac{1}{\widehat{P}(W)} \leqslant 1 + \frac{\overline{\mathrm{nb}}(W^{c})}{\underline{\mathrm{nb}}(W)} = \frac{\underline{\mathrm{nb}}(W) + \overline{\mathrm{nb}}(W^{c})}{\underline{\mathrm{nb}}(W)}$$

therefore

$$\widehat{P}(W) \geqslant \frac{\underline{\mathrm{nb}}(W)}{\underline{\mathrm{nb}}(W) + \overline{\mathrm{nb}}(W^{c})}.$$

In the same manner, since $\widehat{\mathrm{nb}}(W^{c}) \geqslant \underline{\mathrm{nb}}(W^{c})$ and $\widehat{\mathrm{nb}}(W) \leqslant \overline{\mathrm{nb}}(W)$,

$$\frac{1}{\widehat{P}(W)} \geqslant 1 + \frac{\underline{\mathrm{nb}}(W^{c})}{\overline{\mathrm{nb}}(W)} = \frac{\overline{\mathrm{nb}}(W) + \underline{\mathrm{nb}}(W^{c})}{\overline{\mathrm{nb}}(W)}$$

therefore

$$\widehat{P}(W) \leqslant \frac{\overline{\mathrm{nb}}(W)}{\overline{\mathrm{nb}}(W) + \underline{\mathrm{nb}}(W^{c})},$$

which concludes the proof. □

The interval $[\underline{P}(W), \overline{P}(W)]$ can be seen as an imprecise probability measure of W whose imprecision is induced either by the imprecision of the observations [39], or by the imprecision of the counting process.

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

4. Counting transfer

Counting transfer is one of the key features of the QCH technique. It is the main contribution of this paper. The underlying idea of this technique can be expressed as follows. Due to real line partitioning in p cells, a granulation of the number of observations, and therefore of the probability density, occurs. Due to this granulation, one can only estimate the probability of sets of $\Omega \subseteq \mathbb{R}$, the reference interval, corresponding to a union of elementary cells of the partition. Counting transfer provides an estimate of the number of votes purporting any subset of the reference interval.

4.1. Principle of counting transfer

In order to simplify the explanation of counting transfer and its alternatives, we propose to introduce this concept by considering the crisp case.

Let W be a crisp measurable subset (here an interval), $(X_i)_{i=1,\ldots,n}$ be n precise or imprecise observations and $(C_k)_{k=1,\ldots,p}$ be a partition of $\Omega \subseteq \mathbb{R}$ whose granularity is Δ. Let A_k be the imprecise accumulator associated with the kth cell of the partition computed by using expression (18). The information to be estimated is $\mathrm{Nb}(W)$, the number of votes of the set of observations in favor of W. We aim at computing $\widehat{\mathrm{Nb}}(W)$, an imprecise estimate of $\mathrm{Nb}(W)$, by only considering the precise or imprecise accumulators $(A_k)_{k=1,\ldots,p}$.

Fig. 2 shows the crisp uniform partition of reference interval Ω whose counting has to be transferred to W. This transfer is possible only if $W \subseteq \cup_k C_k$, which is the case in this example. The fact that the only information considered is that of the accumulated values associated with each cell of the partition induces ambiguities in the data [5]. In fact, the greater the granularity of the histogram, the more imprecise the histogram-based representation of the underlying pdf. The most favorable case is obtained by fully transferring the count in favor of C_2, C_3 and C_4 onto W. The least favorable case is obtained by fully transferring the count in favor of C_1, C_2, C_4, C_5 and C_6 onto W^c. A median case is obtained by hypothesizing a particular local distribution of votes in each cell. Without any additional information, this distribution is usually supposed to be uniform. In this case, the accumulator value A_k is transferred to W proportionally to $|C_k \cap W|$, i.e. the length (or granularity) of the set $|C_k \cap W|$. First, let us consider accumulations obtained by using precise observations, i.e. the accumulators are precise: $A_k = a_k$.

The first transfer, combining the most favorable and the least favorable cases, is called possibilistic transfer. It provides upper and lower bounds of the pro-W votes count. The lower and upper bounds are given by

$$\mathrm{nb}_N(W) = \sum_{k=1}^{p} a_k N_k(W) \quad \text{(lower)}, \tag{23}$$

$$\mathrm{nb}_\Pi(W) = \sum_{k=1}^{p} a_k \Pi_k(W) \quad \text{(upper)}, \tag{24}$$

with $\Pi_k(W) = \mathrm{Sup}_{x \in C_k} \{\chi_W(x)\}$ and $N_k(W) = 1 - \Pi_k(W^c)$.

Remark 5. If the granularity of W is less than Δ, then $\mathrm{nb}_N(W)$ equals 0 regardless of the values of $(a_k)_{k=1,\ldots,p}$.

The second transfer is called the pignistic transfer. It is based on considering the distribution in each cell as being uniform. Let $f_k(x)$ be the uniform distribution associated with cell C_k:

$$\forall x \in C_k, \quad f_k(x) = p_k = \frac{a_k}{n\Delta}. \tag{25}$$

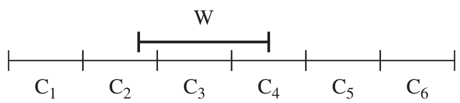


Fig. 2. Binary counting transfer.

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

Since the $(C_k)_{k=1,\ldots,p}$ is a partition of $\mathbb{R}$, we arrive at the naive estimator defined in [36]

$$\forall x \in \Omega, \quad \hat{f}(x) = \sum_{k=1}^{p} f_k(x)\chi_{C_k}(x). \tag{26}$$

Therefore, $\widehat{\mathrm{P}}(W)$, the estimation of the probability of W considering the n observations $(x_i)_{i=1,\ldots,n}$, is given by

$$\begin{aligned}\widehat{\mathrm{P}}(W) &= \int_W \hat{f}(x)\,\mathrm{d}x = \int_W \sum_{k=1}^{p} f_k(x)\chi_{C_k}(x)\,\mathrm{d}x = \sum_{k=1}^{p}\int_W p_k\chi_{C_k}(x)\,\mathrm{d}x \\ &= \sum_{k=1}^{p} p_k \int_W \chi_{C_k}(x)\,\mathrm{d}x = \sum_{k=1}^{p} p_k|W\cap C_k|.\end{aligned} \tag{27}$$

In the same way, $\widehat{\mathrm{nb}}(W)$, an average estimation of the number of observations purporting the subset W, can be given by

$$\widehat{\mathrm{nb}}(W) = n\widehat{\mathrm{P}}(W) = n\sum_{k=1}^{p} p_k|W\cap C_k| = n\sum_{k=1}^{p}\frac{a_k}{n\Delta}|W\cap C_k| = \sum_{k=1}^{p} a_k\frac{|W\cap C_k|}{\Delta} \tag{28}$$

and thus is given by an additive aggregation of the $(a_k)_{k=1,\ldots,p}$ weighted by $|W\cap C_k|/\Delta$.

Now, if the observations are imprecise, then the counting is imprecise. Therefore, the value of a_k has to be replaced by $\underline{a_k}$ in expression (23) and by $\overline{a_k}$ in expression (24). In expression (28), the value of a_k has to be replaced alternatively by $\underline{a_k}$ and $\overline{a_k}$ in order to provide a median lower and a median upper estimate of the votes.

We now present the generalizations of these principles to the fuzzy case (fuzzy partition and imprecise observations). Within our approach, this generalization is based on the α-cut decomposition of the cells of the fuzzy partition [30]. We propose to aggregate the results provided by the crisp approach on each α-level with a Choquet integral.

4.2. Possibilistic counting transfer

Our possibilistic transfer can be seen as a worst-case transfer. This method is a simple generalization of the crisp method presented in Section 4.1. It involves transferring the accumulator a_k to W proportionally to the possibility and necessity of C_k restricted to W. This transfer gives an upper and a lower bound for putative votes of the observations in favor of the imprecise event W. This estimate is an interval.

Definition 1. Let $(C_k)_{k=1,\ldots,p}$ be a partition of $\Omega\subseteq\mathbb{R}$ whose granularity is Δ. Let $A_k = [\underline{a_k}, \overline{a_k}]$ be the imprecise accumulator associated with the kth cell of the partition. $\widehat{\mathrm{Nb}}(W)$, an estimate of the votes of the observations in favor of W by transferring the A_k to W, is given by

$$\widehat{\mathrm{Nb}}(W) = [\underline{\mathrm{nb}}_N(W), \overline{\mathrm{nb}}_\Pi(W)] \tag{29}$$

with

$$\underline{\mathrm{nb}}_N(W) = \sum_{k=1}^{p} \underline{a_k} N_k(W) \quad \text{(lower)}, \tag{30}$$

$$\overline{\mathrm{nb}}_\Pi(W) = \sum_{k=1}^{p} \overline{a_k}\Pi_k(W) \quad \text{(upper)}. \tag{31}$$

Property 4. The new upper and lower estimates (30) and (31) are simple Choquet integrals [22] over the different level cuts of upper and lower estimates given by expressions (23) and (24).

Proof. Let $W^\alpha = \{x, \mu_W(x)\geqslant\alpha\}$ be the α-level cut of W, and C_k^α be the α-level cut of C_k. For each level α, $\overline{\mathrm{nb}}_\Pi^\alpha(W) = \sum_{k=1}^{p}\overline{a_k}\Pi_k(W^\alpha)$, with $\Pi_k(W^\alpha) = \mathrm{Sup}_{x\in\Omega}\{\min(\chi_{C_k^\alpha}(x), \chi_{W^\alpha}(x))\}$. The integral of all the $\overline{\mathrm{nb}}_\Pi^\alpha(W)$ is

$\int_0^1 \sum_{k=1}^{p} \overline{a_k} \Pi_k(W^\alpha)\,\mathrm{d}\alpha$. Let $A^\alpha = \{k, \Pi_k(W^\alpha) = 1\}$, since W and C_k are fuzzy subsets, the level subsets are nested, i.e. for any $\mathrm{d}\alpha \in [0, 1-\alpha]$, $W^\alpha \subseteq W^{\alpha+\mathrm{d}\alpha}$ and $C_k^\alpha \subseteq C_k^{\alpha+\mathrm{d}\alpha}$ and therefore $A^\alpha \subseteq A^{\alpha+\mathrm{d}\alpha}$. Let $\alpha_k = \Pi_k(W)$, we can suppose, without any loss of generality, that the α_k are sorted in increasing order, i.e. $\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_p \leqslant \alpha_{p+1}$, with $\alpha_{p+1} = 1$ by convention. Then $A^{\alpha_k} = \{1, \ldots, k\}$. Let $n(\alpha)$ be the index such that $\alpha_{n(\alpha)} \leqslant \alpha < \alpha_{n(\alpha)+1}$, thus $\sum_{k=1}^{p} \overline{a_k} \Pi_k(W^\alpha) = \sum_{k=1}^{n(\alpha)} \overline{a_k}$. Therefore, with the α-levels being discrete, the continuous integral reduces to a discrete integral: $\int_0^1 \sum_{k=1}^{p} \overline{a_k} \Pi_k(W^\alpha)\,\mathrm{d}\alpha = \sum_{i=1}^{p} \sum_{k=1}^{n(\alpha_i)} \overline{a_k}(\alpha_{k+1} - \alpha_k)$. By definition, $n(\alpha_i) = i$, then

$$\overline{\mathrm{nb}}_\Pi(W) = \int_0^1 \sum_{k=1}^{p} \overline{a_k} \Pi_k(W^\alpha)\,\mathrm{d}\alpha = \sum_{i=1}^{p} \sum_{k=1}^{i} \overline{a_k}(\alpha_{k+1} - \alpha_k) = \sum_{i=1}^{p} \overline{a_k}\alpha_k = \sum_{k=1}^{p} \overline{a_k} \Pi(W).$$

Deriving expression (30) from expression (23) can be simply achieved by using complementation. First note that $N_k(W) = 1 - \Pi_k(W^c)$ and $N_k(W^\alpha) = 1 - \Pi_k((W^\alpha)^c)$. Thus

$$\begin{aligned}\int_0^1 \sum_{k=1}^{p} \underline{a_k} N_k(W^\alpha)\,\mathrm{d}\alpha &= \int_0^1 \sum_{k=1}^{p} \underline{a_k}(1 - \Pi_k(W^c))\,\mathrm{d}\alpha \\ &= \int_0^1 \sum_{k=1}^{p} \underline{a_k}\,\mathrm{d}\alpha - \int_0^1 \sum_{k=1}^{p} \underline{a_k} \Pi_k((W^\alpha)^c)\,\mathrm{d}\alpha \\ &= \sum_{k=1}^{p} \underline{a_k} - \sum_{k=1}^{p} \underline{a_k} \Pi_k(W^c) = \sum_{k=1}^{p} \underline{a_k}(1 - \Pi_k(W^c)) = \sum_{k=1}^{p} \underline{a_k} N_k(W). \qquad \square\end{aligned}$$

Property 5. If the observations are precise, then $\overline{\mathrm{nb}}_\Pi(\Omega) = \underline{\mathrm{nb}}_N(\Omega) = n$.

Proof. Since $\forall k \in \{l, \ldots, p\}$, $\Pi_k(\Omega) = N_k(\Omega) = 1$, $\overline{\mathrm{nb}}_\Pi(\Omega) = \sum_{k=1}^{p} a_k = n$ and

$$\underline{\mathrm{nb}}_N(\Omega) = \sum_{k=1}^{p} a_k = n.$$

Property 6. If the observations are imprecise, then $\underline{\mathrm{nb}}_N(\Omega) \leqslant n \leqslant \underline{\mathrm{nb}}_\Pi(\Omega)$.

This property can be trivially proved in the same manner as Property 2.

Remark 6. The lower transfer given by expression (30) vanishes if $|W|$ (the granularity of W) is lower than Δ (the granularity of the histogram) because the transfer is no longer normalized:

$$\sum_{k=1}^{p} N(W; C_k) < 1 \quad \text{if } |W| < \Delta. \tag{32}$$

Remark 7. Compared to the crisp case, the property $\underline{\mathrm{nb}}_N(W) \leqslant \underline{\mathrm{nb}}(W) \leqslant \overline{\mathrm{nb}}(W) \leqslant \overline{\mathrm{nb}}_\Pi(W)$ where $\underline{\mathrm{nb}}(W) \sum_{i=1}^{n} N_i(W)$ and $\overline{\mathrm{nb}}(W) \sum_{i=1}^{n} \Pi_i(W)$ is generally lost. This is mainly due to the Choquet-based aggregation.

4.3. Precise pignistic counting transfer

The precise pignistic transfer is inspired by the transferable belief model proposed by Smets [38]. It consists of transferring the value of a_k to W proportionally to the coverage of W and C_k. A similar counting transfer has been used to prove the convergence of fpbh in mean squared error [21] and integrated mean square error [47].

If observations are precise, then accumulators are precise. The Choquet integral of expression (28) gives

$$\widehat{\mathrm{nb}}(W) = \int_0^1 \sum_{k=1}^{p} a_k \frac{|W^\alpha \cap C_k^\alpha|}{\Delta}\,\mathrm{d}\alpha = \sum_{k=1}^{p} \frac{a_k}{\Delta} \int_0^1 |W^\alpha \cap C_k^\alpha|\,\mathrm{d}\alpha = \sum_{k=1}^{p} a_k \frac{|W \cap C_k|}{\Delta}, \tag{33}$$

where the intersection of two fuzzy subsets is defined by the T-norm min. In this case, expression (33) is equivalent to expression (28). This additive way of aggregating the $(a_k)_{k=1,\ldots,p}$ consists of defining, for each cell C_k, a probability $\rho_w(k)$ for a_k to be a value to consider for an estimation of the votes purporting W:

$$\rho_w(k) = \frac{|W \cap C_k|}{|W|}. \tag{34}$$

When observations are imprecise, then the estimation of the votes purporting W is an interval:

$$\widehat{\mathrm{Nb}}(W) = [\underline{\widehat{\mathrm{nb}}}(W), \overline{\widehat{\mathrm{nb}}}(W)] \tag{35}$$

with

$$\underline{\widehat{\mathrm{nb}}}(W) = \frac{|W|}{\varDelta} \sum_{k=1}^{p} \rho_W(k) \underline{a_k} \quad \text{and} \quad \overline{\widehat{\mathrm{nb}}}(W) = \frac{|W|}{\varDelta} \sum_{k=1}^{p} \rho_W(k) \overline{a_K}. \tag{36}$$

Two important assumptions were put forward to provide this median transfer. First, the count in each cell is supposed to be uniform for any α-level set (which is the principle of the pignistic transfer). Second, the cells are supposed to be non-interactive fuzzy numbers. In fact, the second hypothesis is violated by construction because the α-level sets of two consecutive cells have a non-empty intersection when $\alpha < 0.5$.

In most studies based on an fpbh, a pignistic-like method is used for transferring knowledge about the densities $p_k = a_k/(n\varDelta)$ associated with each cell C_k onto any values of Ω. This transfer simply consists of interpolating these densities with weights $\rho(k)$ obtained by convolving kernels of the partition with another kernel centered on the value on which the underlying pdf has to be estimated. The former kernel represents the hypothesized underlying distribution of votes in each cell. This approach coincides with the pignistic approach if W is a usual crisp subset and the accumulators are precise:

$$\widehat{\mathrm{nb}}(W) = \sum_{k=1}^{p} \frac{a_k}{\varDelta} |W \cap C_k| = \sum_{k=1}^{p} \frac{a_k}{\varDelta} \int_W \mu_{C_k}(x)\,\mathrm{d}x. \tag{37}$$

Let $\eta(x) = \chi_W(x)/|W|$ be the uniform kernel associated with W, then

$$\widehat{\mathrm{nb}}(W) = \Gamma \sum_{k=1}^{p} \frac{a_k}{\varDelta} \rho_W(k) \quad \text{with } \rho_W(k) = \int_\Omega \mu_{C_k}(x)\eta(x)\,\mathrm{d}x. \tag{38}$$

The estimation $\widehat{P}(W)$ of the probability W is then given by

$$\widehat{\mathrm{P}}(W) = \frac{\widehat{\mathrm{nb}}(W)}{n} = |W| \sum_{k=1}^{p} \frac{a_k}{n\varDelta} \rho_W(k), \tag{39}$$

which is the interpolating method proposed in [20,21,32,35,44].

Note that using the precise pignistic transfer to estimate the votes purporting a crisp subset W of Ω coincides with estimating these votes via a Parzen–Rosenblatt kernel density estimation.

Property 7. Let $(C_k)_{k=1,\ldots,p}$ be a strong fuzzy partition of $\Omega \subseteq \mathbb{R}$, then there is a summative kernel κ such that $\widehat{\mathrm{nb}}(W) = n \int_W \hat{f}_\kappa(x)\,\mathrm{d}x$.

Proof. For each k, $a_k = \sum_{i=1}^{n} \mu_{C_k}(x_i)$. Since $(C_k)_{k=1,\ldots,p}$ is a strong partition of Ω, $\forall x \in \Omega$, $\sum_{k=1}^{p} \mu_{C_k}(x) = 1$ and, due to Property 1, $n = \sum_{k=1}^{p} a_k$. Let E be the basic kernel such that $\forall x \in \Omega$, $\mu_{C_k}(x) = \mu_E(x - m_k)$, then

$$\begin{aligned}\widehat{\mathrm{nb}}(W) &= \sum_{k=1}^{p} a_k \left(\int_W \frac{\mu_{C_k}(x)}{\varDelta}\,\mathrm{d}x \right) = \sum_{k=1}^{p} \left(\sum_{i=1}^{n} \mu_{C_k}(x_i) \right) \left(\int_W \frac{\mu_{C_k}(x)}{\varDelta}\,\mathrm{d}x \right) \\ &= \sum_{k=1}^{p} \left(\sum_{i=1}^{n} \left(\int_W \frac{\mu_{C_k}(x)}{\varDelta} \mu_{C_k}(x_i)\,\mathrm{d}x \right) \right) = \int_W \left(\sum_{i=1}^{n} \left(\sum_{k=1}^{p} \left(\frac{\mu_{C_k}(x)}{\varDelta} \mu_{C_k}(x_i) \right) \right) \right) \mathrm{d}x.\end{aligned}$$

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

Now let

$$\kappa_i(x-x_i)=\sum_{k=1}^{p}\left(\frac{\mu_{C_k}(x)}{\Delta}\mu_{C_k}(x_i)\right).$$

κ_i is a summative kernel since $\forall_i \in \{1,\cdots,n\}$,

$$\int_\Omega \kappa_i(x-x_i)\,\mathrm{d}x=\int_\Omega\left(\sum_{k=1}^{p}\left(\frac{\mu_{C_k}(x)}{\Delta}\mu_{C_k}(x_i)\right)\right)\mathrm{d}x=\sum_{k=1}^{p}\int_\Omega\frac{\mu_E(x-m_k)}{\Delta}\mu_{C_k}(x_i)\,\mathrm{d}x$$

$$=\sum_{k=1}^{p}\mu_{C_k}(x_i)\int_\Omega\frac{\mu_E(x-m_k)}{\Delta}\,\mathrm{d}x=\sum_{k=1}^{p}\mu_{C_k}(x_i)\frac{1}{\Delta}\int_\Omega\mu_E(x-m_k)\,\mathrm{d}x=\sum_{k=1}^{p}\mu_{C_k}(x_i)\frac{1}{\Delta}\Delta=1$$

then

$$\widehat{\mathrm{nb}}(W;(x_i))=\int_W\left(\sum_{i=1}^{n}\kappa_i(x-x_i)\right)\mathrm{d}x=n\int_W\left(\frac{1}{n}\sum_{i=1}^{n}\kappa_i(x-x_i)\right)\mathrm{d}x=n\int_W\hat{f}_\kappa(x)\,\mathrm{d}x,$$

with

$$\hat{f}_\kappa(x)=\frac{1}{n}\sum_{i=1}^{n}\kappa_i(x-x_i).\qquad\square$$

4.4. Imprecise pignistic counting transfer

Our second proposition for pignistic transfer aims to account for the interactivity of two consecutive cells in order to propagate on $\widehat{\mathrm{Nb}}(W)$ the counting imprecision due to granulation. Let us go back to α-level sets of the fuzzy partition.

Fig. 3a shows a crisp interval W and the three cells of the partition interacting with W. When level α is lower than 0.5, then the $(C_k^\alpha)_{k=1,\ldots,p}$ have non-empty intersections (C_k^α is the α-level set of C_k). Let us consider the two cells C_1 and C_2. Since $C_1^\alpha\cap C_2^\alpha\neq\emptyset$, the counting density in the interval $C_1^\alpha\cap C_2^\alpha$ is ill-known (Fig. 3c). The precise pignistic transfer presented previously assumes that the local counting density in $C_1^\alpha\cap C_2^\alpha$ can be obtained by averaging a_1 and a_2. Imprecise pignistic transfer considers this density as imprecise and proportional to the most specific interval including a_1 and a_2, i.e. $[\min(a_1,a_2),\max(a_1,a_2)]$. Thus the lower contribution of the two cells C_1 and C_2 to the pignistic transfer towards W can be obtained by

$$|W\cap C_1^\alpha\cap C_2^\alpha|\min(a_1,a_2)+|W\cap B_1^\alpha|a_1+|W\cap B_2^\alpha|a_2,$$

while the upper contribution is given by

$$|W\cap C_1^\alpha\cap C_2^\alpha|\max(a_1,a_2)+|W\cap B_1^\alpha|a_1+|W\cap B_2^\alpha|a_2,$$

where $B_1^\alpha=C_1^\alpha\backslash(C_1^\alpha\cap C_2^\alpha)$ and $B_2^\alpha=C_2^\alpha\backslash(C_1^\alpha\cap C_2^\alpha)$. This kind of transfer therefore impacts the ambiguity in counting due to fuzzy partitioning on the estimated transferred value. When level α is greater than 0.5, then the $(C_k^\alpha)_{k=1,\ldots,p}$ have empty intersections and the pignistic transfer is precise.

The imprecise pignistic transfer gives an imprecise estimate of $\mathrm{nb}(W)$ regardless of the precision of the data. It is computed by a discrete Choquet integral [14] and requires a complete sorting of the $(\overline{a_k})_{k=1,\ldots,p}$ and the $(\underline{a_k})_{k=1,\ldots,p}$.

Definition 2. Let $(C_k)_{k=1,\ldots,p}$ be a partition of $\mathbb{R}$ whose granularity is Δ. Let $A_k=[\underline{a_k},\overline{a_k}]$ be the imprecise accumulator associated with the kth cell of the partition. $\widehat{\overline{\mathrm{nb}}}(W)$, an upper estimate of the votes of the observations in favor of W by transferring the A_k to W is given by

$$\widehat{\overline{\mathrm{nb}}}(W)=\frac{|W|}{\Delta}\sum_{k=1}^{p}\overline{a_{(k)}}(\upsilon_W(S_{(k)})-\upsilon_W(S_{(k+1)})),\tag{40}$$

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

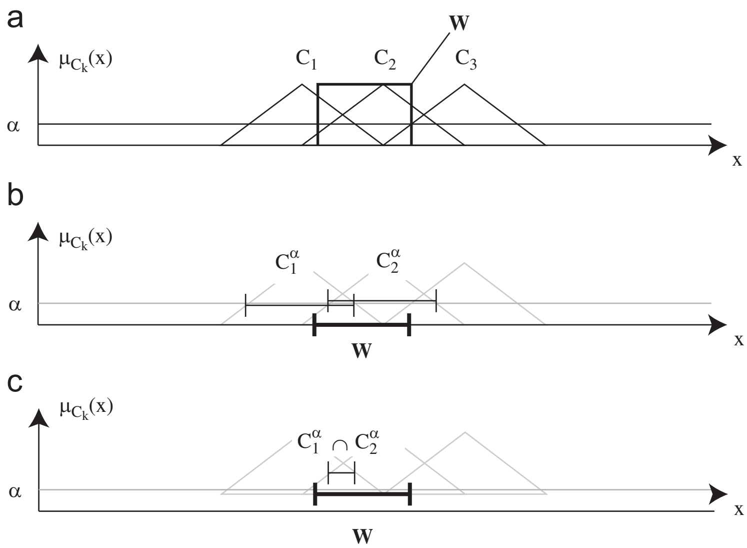


Fig. 3. α-Level sets for imprecise pignistic transfer.

where $(\cdot)$ indicates a permutation sorting the upper accumulators such that $\overline{a_{(1)}} \leqslant \overline{a_{(2)}} \leqslant \cdots \leqslant \overline{a_{(p)}}$, $S_{(k)}$ is a coalition (group of cells) defined by

$$S_{(k)} = \{(k), \ldots, (p)\} \quad \text{and} \quad S_{(p+1)} = \emptyset \tag{41}$$

and $\upsilon_W(S_{(k)})$ is a confidence measure on this coalition defined by

$$\upsilon_W(S) = \xi_W |E_S \cap W|, \tag{42}$$

with $E_S = \bigcup_{i \in S} C_i$ and ξ_W a normalization factor:

$$\xi_W = \frac{1}{|(C_1 \cup \cdots \cup C_p) \cap W|}, \tag{43}$$

the intersection being defined by the T-norm min and the union by the T-conorm max.

Definition 3. Let $(C_k)_{k=1,\ldots,p}$ be a partition of $\Omega \subseteq \mathbb{R}$ whose granularity is Δ. Let $A_k = [\underline{a_k}, \overline{a_k}]$ be the imprecise accumulator associated with the kth cell of the partition. $\widehat{\underline{\mathrm{nb}}}(W)$, a lower estimate of the votes of the observations in favor of W by transferring the A_k to W is given by

$$\widehat{\underline{\mathrm{nb}}}(W) = \frac{|W|}{\Delta} \sum_{k=0}^{p} \underline{a_{(k)}}(\upsilon_W(T_{(k)}) - \upsilon_W(T_{(k+1)})), \tag{44}$$

where $(\cdot)$ indicates a permutation sorting the lower accumulators such that: $\underline{a_{(1)}} \geqslant \underline{a_{(2)}} \geqslant \cdots \geqslant \underline{a_{(p)}}$, $T_{(k)}$ is also a coalition but based on sorting of the $\underline{a_k}$ in decreasing order:

$$T_{(k)} = \{(k), \ldots, (p)\} \quad \text{and} \quad T_{(p+1)} = \emptyset. \tag{45}$$

Property 8. Let $(C_k)_{k=1,\ldots,p}$ be a strong partition of $\Omega \subseteq \mathbb{R}$, the confidence measure υ_W is submodular (concave): $\forall (S,T) \subseteq \{1, \ldots, p\}^2$, $\upsilon_W(S) + \upsilon_W(T) \geqslant \upsilon_W(S \cup T) + \upsilon_W(S \cap T)$.

Proof. Let S and T be two subsets of $\{1, \ldots, p\}$. First, note that $E_{S\cap T} \subseteq E_S \cap E_T$: if $k \in S \cap T$ then $k \in S$ and $k \in T$ therefore $C_k \subseteq E_S$ and $C_k \subseteq E_T$ thus $C_k \subseteq E_S \cap E_T$ and therefore $E_{S\cap T} \subseteq E_S \cap E_T$. The converse inclusion is false. A counter example is given by considering $S = \{k\}$ and $T = \{k+1\}$. Since two contiguous cells intersect $E_S \cap E_T = C_k \cap C_{k+1} \neq \emptyset$ but $E_{S\cap T} = E_\emptyset = \emptyset$. Secondly, note that $E_{S\cup T} = E_S \cup E_T$: $\forall k \in S \cup T$ then $k \in S$ or $k \in T$ therefore $C_k \subseteq E_S$ or $C_k \subseteq E_T$ thus $C_k \subseteq E_S \cup E_T$ and therefore $E_{S\cup T} \subseteq E_S \cup E_T$. Conversely, $E_S \cup E_T \not\subset E_{S\cup T}$ means that $\exists k \in \{1, \ldots, p\}$ such that $C_k \subseteq E_S \cup E_T$ and $k \notin S$ and $k \notin T$, which is in complete contradiction with the fact that $(C_k)_{k=1,\ldots,p}$ is a strong fuzzy partition (expression (13)). Thus $E_{S\cup T} \subseteq E_S \cup E_T$ and $E_S \cup E_T \subseteq E_{S\cup T}$ therefore $E_{S\cup T} = E_S \cup E_T$. Since $|E_S| + |E_T| = |E_S \cup E_T| + |E_S \cap E_T|$ with $E_{S\cup T} = E_S \cup E_T$ and $E_{S\cap T} \subseteq E_S \cap E_T$ then $|E_S \cup E_T| = |E_{S\cup T}|$ and $|E_S \cap E_T| \geqslant |E_{S\cap T}|$ and therefore $|E_S| + |E_T| \geqslant |E_{S\cup T}| + |E_{S\cap T}|$. Since $\forall S \subseteq \{1, \ldots, p\}$, $\upsilon_W(S) = \xi_W |E_S \cap W|$, thus $|E_{S\cup T}| + |E_{S\cap T}| \leqslant |E_S| + |E_T|$ implies $\upsilon_W(S \cup T) + \upsilon_W(S \cap T) \leqslant \upsilon_W(S) + \upsilon_W(T)$ which completes the proof. $\square$

Since the confidence measure υ_W is concave, it defines a set of probabilities denoted $\Im(\upsilon_W)$ called the core of υ_W [8], satisfying

$$\forall P \in \Im(\upsilon_W), \quad \forall S \subseteq \{1, \ldots, p\}, \quad \upsilon_W^c(S) \leqslant P(S) \leqslant \upsilon_W(S). \tag{46}$$

P is said to be dominated by υ_w. A useful property from the work of Dennenberg [8] and Schmeidler [34] is related to the consequences of the domination of concave capacities: the interval provided by aggregating the $(A_k)_{k=1,\ldots,p}$ with the submodular capacity υ_W is the smallest interval that encloses any precise estimation obtained by aggregating the $(A_k)_{k=1,\ldots,p}$ with any probability dominated by υ_W. Therefore, the length of this interval evaluates the extent of variability of all usual precise interpolations performed with any set of probabilistic weights induced by a probability dominated by υ_W.

Remark 8. If $(C_k)_{k=1,\ldots,p}$ is a crisp partition, then υ_W is additive and thus the imprecise pignistic transfer becomes a precise pignistic transfer.

We conjecture the two following properties:

Conjecture 1. *When associating an fpbh with imprecise counting transfer, the interval-valued probability estimation given by expressions* (20) *and* (21) *is a good measure of the variability induced by using different summative kernel shapes whose bandwidths have been adapted to the granularity of the histogram* (*see* [19,36] *for bandwidth adaptation*).

Conjecture 2. *Let* $(C_k)_{k=1,\ldots,p}$ *be a strong fuzzy partition of* Ω *and* W *be a subset of* Ω, *the imprecise transfer capacity* υ_W *dominates the precise transfer probability induced by* ρ_W, *i.e.*, $\forall S \subseteq \{1, \ldots, p\}$,

$$\upsilon_W(S) \geqslant P_W(S) = \sum_{k\in S} \rho_W(k).$$

This conjecture is easy to prove on singletons.

Property 9. $\forall k \in \{1, \ldots, p\}$, $\upsilon_W(\{k\}) \geqslant \rho_W(k)$.

Proof. Since $(C_k)_{k=1,\ldots,p}$ is a strong fuzzy partition of Ω, $\mathrm{Sup}_{k\in\{1,\ldots,p\}} \mu_{C_k}(x) \leqslant 1 = \sum_{k=1}^{p} \mu_{C_k}(x)$. Therefore $|(C_1 \cup \cdots \cup C_p) \cap W| = \int_W \mathrm{Sup}_{k\in\{1,\ldots,p\}} \mu_{C_k}(x)\,\mathrm{d}x \leqslant \int_W \sum_{k=1}^{p} \mu_{C_k}(x)\,\mathrm{d}x = \sum_{k=1}^{p} |C_k \cap W|$. Thus

$$\upsilon_W(\{k\}) = \frac{|C_k \cap W|}{|(C_1 \cup \cdots \cup C_p) \cap W|} \geqslant \frac{|C_k \cap W|}{\sum_{k=1}^{p} |C_k \cap W|} = \rho_W(k). \qquad \square$$

If Conjecture 1 is proved, then the interval-valued relative frequency provided by an imprecise counting transfer will have the special meaning of *the set of all relative frequencies that would have been obtained by a kernel-density estimation with a class of adapted kernels with various shapes*. The proof of Conjecture 2 would enable us to consider imprecise pignistic transfer as a way of propagating the imprecision induced by the partition to the imprecision of the computed relative frequencies, i.e. to the estimate of the probability of a measurable set W of $\Omega \subseteq \mathbb{R}$.

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

Property 10. If the intersection and the union in expressions (42) and (43) are defined by the Lukasiewicz operators, then the imprecise pignistic transfer defined by (40) and (44) is equivalent to the precise pignistic transfer defined by (33), the intersection being defined by the T-norm min.

Proof. Let $S \subseteq \{1, \dots, p\}$. Let us remark that, considering the Lukasiewicz operators, $\mu_{E_S \cap W}(x) = \max(0, \chi_W(x) + \mu_{E_S}(x) - 1)$ with $\mu_{E_S}(x) = \max(1, \sum_{i \in S} \mu_{C_i}(x)) = \sum_{i \in S} \mu_{C_i}(x)$. Therefore, if $x \in W$, $\mu_{E_S \cap W}(x) = \sum_{i \in S} \mu_{C_i}(x)$, else $\mu_{E_S \cap W}(x) = 0$. Thus $\forall x \in \Omega$, $\mu_{C_1 \cup \dots \cup C_p}(x) = \max(1, \sum_{k=1}^{p} \mu_{C_k}(x)) = 1$ and therefore $\xi_W = 1/|W|$. Now, let $k \in \{1, \dots, p\}$ and $S_{(k)}$ and $S_{(k+1)}$ be two subsets defined by expression (41): $S_{(k)} = S_{(k+1)} \cup \{(k)\}$. Thus $\sum_{i \in S_{(k)}} \mu_{C_i}(x) = (\sum_{i \in S_{(k+1)}} \mu_{C_i}(x)) + \mu_{C_{(k)}}(x)$ and therefore $\upsilon_W(S_{(k)}) - \upsilon_W(S_{(k+1)}) = (1/|W|) \int_W \mu_{C_{(k)}}(x)$. Since $\int_W \mu_{C_{(k)}}(x) = |C_{(k)} \cap W|$ then $|W|/\Delta \sum_{k=1}^{p} a_{(k)}(\upsilon_W(S_{(k)}) - \upsilon_W(S_{(k+1)})) = \sum_{k=1}^{p} a_{(k)}(|C_{(k)} \cap W|/\Delta)$, which completes the proof. □

5. Counting transfer experiments

In this section, we propose some experiments to illustrate the different properties of counting transfer. We will consider the classic Old Faithful Geyser dataset that has been extensively used by Silverman [36]. It involves 107 observations of the duration of eruptions of the Old Faithful Geyser in Yellowstone National Park. The underlying density is bimodal. The precision of the observed data is unknown. However, for our experiments, we will suppose that the observation's precision is identically distributed and that it can be estimated by the median of the least absolute time between two different observations. This value equals 0.025 min.

To perform the experiments, we divided the reference interval $\Omega = [0, 6]$ min into 1000 equally spaced samples $(\omega_i)_{i=1,\dots,1000}$. At each location ω_i, an interval W_i has been defined. All intervals have the same granularity equal to twice the granularity of the histogram (to avoid a weak possibilistic lower transfer). The experiments involve comparing different ways of estimating $P(W_i)_{i=1,\dots,1000}$ based on the observations of eruption duration.

5.1. Counting transfer and density estimation

This section aims at comparing the Parzen–Rosenblatt kernel-based density estimation method with the QCH-based density estimation method. To achieve this comparison, we perform a kernel density estimation at each location ω_i by using the nine different bounded and non-bounded kernels whose basic functions are given in Table 1. From Table 1, we can retrieve a kernel κ_∇ with bandwidth ∇ by $\kappa_\nabla(x) = (1/\nabla)\kappa(x/\nabla) \cdot \chi_{[-1,1]}$ is the characteristic function of the interval $[-1, 1]$. Let κ_∇ be the considered kernel, the $P_{\kappa_\nabla}(W_i)$ value is obtained by discrete integration of $\hat{f}_{\kappa_\nabla}(\omega_i)$: $P(W_i) \approx \sum_{j=1}^{1000} \hat{f}_{\kappa_\nabla}(\omega_j)\chi_{W_i}(\omega_j)$.

For the QCH part of the experiment, the fpbh is built on a triangular partition with granularity Δ of the reference interval Ω. Then we estimate the number of observations purporting each crisp interval W_i by different transfer methods (see Section 4). This estimation is then normalized by using expressions (20) and (21). In this experiment, the different summative kernels were adapted to the partition granularity [19].

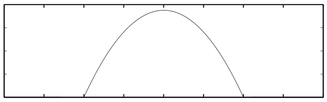
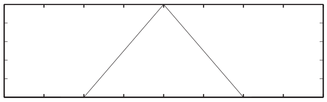
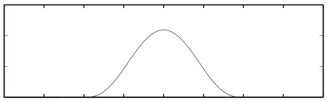
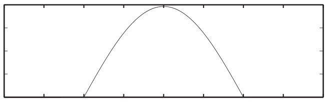
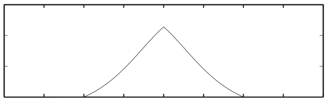
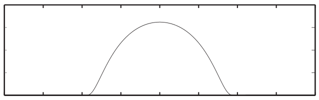
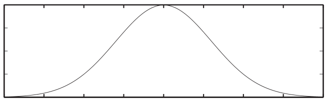
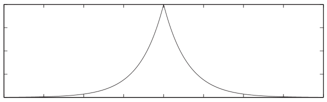
Figs. 4–6 show the estimates of $P(W_i)$ by transferring, on each W_i, the accumulated values of an fpbh with granularity $\Delta = 0.3$ min. Fig. 4 shows the precise pignistic transfer, Fig. 5 shows the imprecise pignistic transfer, Fig. 6 shows the possibilistic transfer. In each of these figures, the density estimates given by using the nine unimodal bounded and non-bounded summative kernels are plotted in cyan, while the QCH estimate is plotted in black for median, blue for upper, and red for lower estimates.

Note, in Fig. 5, that the interval provided by the imprecise pignistic transfer seems to be representative of variations in density estimation when the kernel varies in the class of positive summative unimodal kernels whose granularity is equal to Δ. This is Conjecture 1. In the same way, the pessimistic estimate provided by the possibilistic transfer bounds the whole set of estimates provided by this class of kernels (Fig. 6), except for areas where the local observation density is too low to ensure convergence of any kernel estimation towards the underlying density estimate. Finally, Fig. 4 illustrates Property 7, i.e. the density estimate provided by the pignistic transfer is equivalent to a Parzen–Rozenblatt density estimation with an adapted granularity.

A relevant observation can be made by simply comparing Figs. 5 and 6: both possibilistic and imprecise pignistic transfer seem to partially include the variability in density estimation due to variation in kernel shapes, but in a more specific way for imprecise pignistic transfer. This is mostly due to the fact that the chosen bandwidth is wide enough

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

Table 1
Nine basic summative kernels.

Kernel name	Kernel expression	Kernel shape
Epanechnikov	$\kappa(u) = \frac{3}{4}(1-u^2)\chi_{[-1,1]}(u)$	
Triangular	$\kappa(u) = (1-\lvert u\rvert)\chi_{[-1,1]}(u)$	
Triweight	$\kappa(u) = \frac{35}{32}(1-u^2)^3\chi_{[-1,1]}(u)$	
Cosine	$\kappa(u) = \dfrac{\Pi}{4}\cos\left(\dfrac{\Pi}{2}u\right)\chi_{[-1,1]}(u)$	
Uniform	$\kappa(u) = \frac{1}{2}\chi_{[-1,1]}(u)$	
Ratio	$\kappa(u) = \dfrac{1}{\dfrac{\pi}{2}-\ln(2)}\left(\dfrac{1-\lvert u\rvert}{1+u^2}\right)\chi_{[-1,1]}(u)$	
Smooth	$\kappa(u) = 0.44\exp\left(\dfrac{1}{u^2-1}\right)\chi_{[-1,1]}(u)$	
Gaussian	$\kappa(u) = \dfrac{1}{\sqrt{2\Pi}}\exp\left(-\dfrac{u^2}{2}\right)$	
Exponential	$\kappa(u) = \frac{1}{2}\exp(-\lvert u\rvert)$	

to ensure appropriate smoothing. When the bandwidth is too narrow, the imprecise pignistic transfer does not provide a good bracketing of the density estimation variability while the possibilistic transfer still does. Fig. 7 illustrates this bad bracketing. The granularity of the kernels is reduced to 0.006 min. In Fig. 7a the nine classic density estimates are outside the bounds of the imprecise density estimates, while in Fig. 7b, the possibilistic-based approach still brackets the nine density estimates.

5.2. Counting transfer and fast computation

Computation of the density estimate provided by QCH is faster than conventional density estimation. This fact is due to the pre-classification performed by the histogram [41]. Let us define, as a time unit, the computation time of the interaction between a kernel and an observation (computation of $\kappa_{\nabla}(x)$). Let n be the number of observations, q be the number of points on the real line on which the density estimate has to be computed and p the number of QCH cells. Kernel density estimation requires $q \cdot n$ elementary operations while only $p \cdot n$ elementary operations are needed to compute the histogram and $p \cdot q$ elementary operations are needed to compute the density estimate, i.e. $p(q+n)$ elementary operations. Since usually $p \ll q$ and $p \ll n$ then $p(q+n) \ll q \cdot n$, i.e. the computation time is reduced. To

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

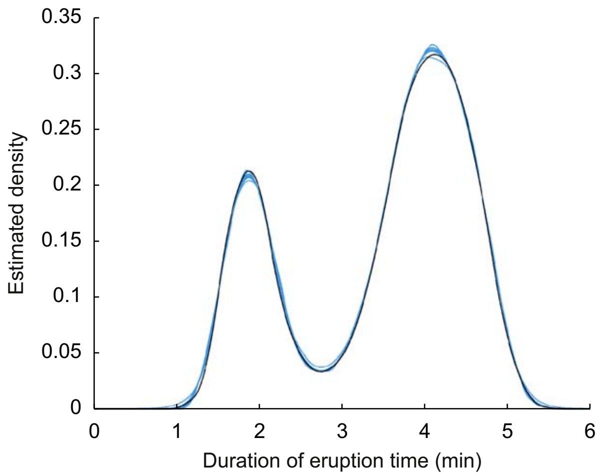


Fig. 4. Comparison of the QCH estimation using precise pignistic transfer (in black) with nine classic density estimates with the same granularity (in light blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

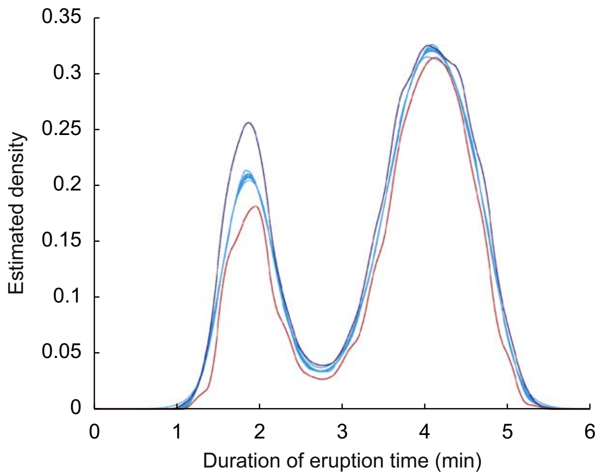


Fig. 5. Comparison of the QCH estimation using imprecise pignistic transfer (blue—upper, red—lower) with nine classic density estimates with the same granularity (cyan). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

illustrate this property, we plotted in Fig. 8 the ratio of the computation time of the classic density estimate against the computation time of the density estimate with QCH. The histogram is made of $p = 30$ cells, the density is estimated on $q = 1000$ points and the number of observations is increased (artificially by bootstrapping) from 100 to 10 000. In both cases, a triangular kernel has been chosen. The plot in Fig. 8 has the expected shape (with a 10th supplementary factor in favor of the QCH method).

5.3. Counting transfer and histogram granularity

This experiment aims at illustrating the ability of imprecise pignistic transfer to impact the ambiguities in data due to the (fuzzy) partitioning of Ω to the estimated pdf. Fig. 9 plots the density estimate obtained by precise (black) and imprecise (red—lower, blue—upper) pignistic transfer for different partition granularities. Clearly the interval-valued estimated density collapses with the partition granularity. The interval provided by the imprecise pignistic transfer is

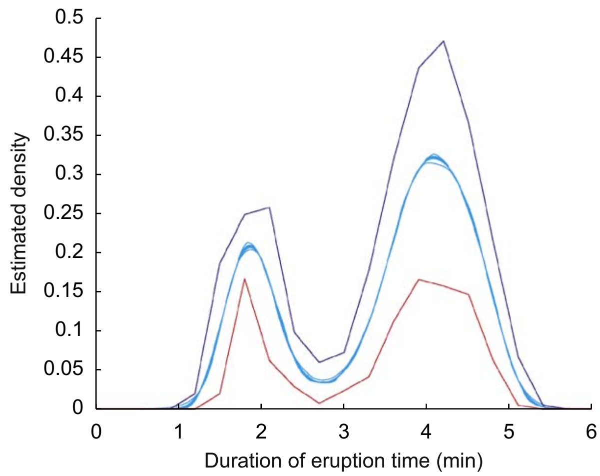


Fig. 6. Comparison of the QCH estimation using the possibilistic transfer (blue—upper, red—lower) with nine classic density estimates with the same granularity (cyan). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

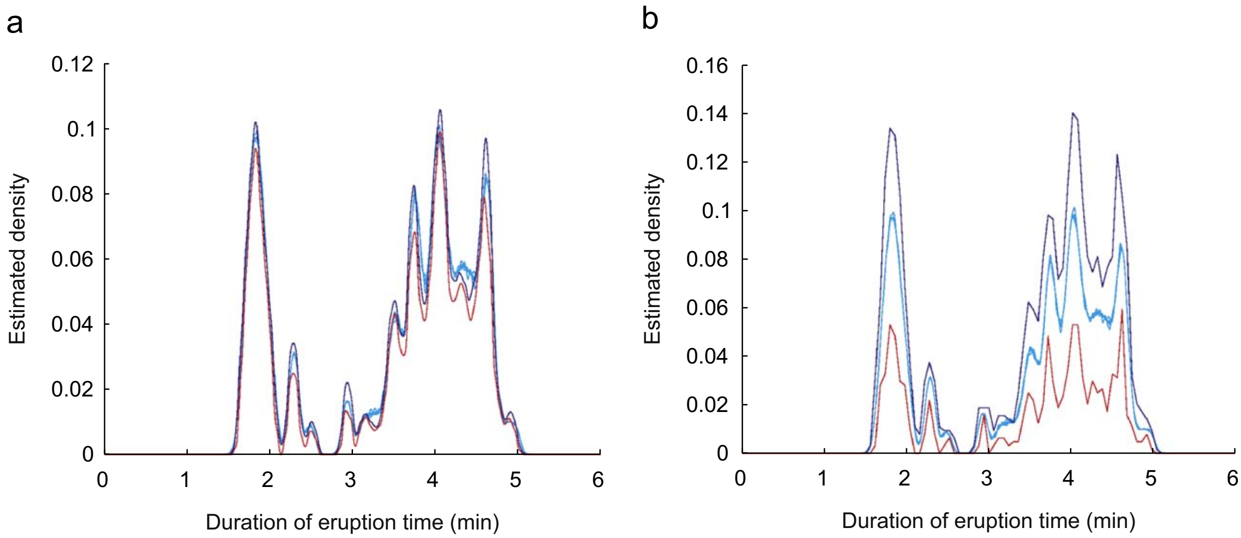


Fig. 7. Comparison of the QCH estimation using imprecise pignistic (a) and possibilistic (b) transfer superimposed on nine classic density estimates with the same granularity with $\Delta = 0.06$ min.

a signature of the granulation induced by the partition. Moreover, this experiment illustrates Conjecture 2 since the precise estimated density is always included in the interval-valued imprecise estimated density.

Although using a QCH lowers the influence of partitioning on the reconstructed density, this influence still exists. In fact, when considering the proof of Property 7, this dependence clearly appears since the shape of the equivalent kernel depends on the relative position of the observation to modes of the partition. We illustrate this dependence by translating the positions of histogram partitions, and estimating the density at the same locations. Fig. 10a plots the precise estimations of density for five different positions, superimposed on one imprecise estimation of the density for the first position. Every precise estimation is included in the interval-valued imprecise estimation. Therefore, the imprecision of the interval-valued imprecise estimation seems to be a good measure of the partitioning influence on the estimated density. Fig. 10b plots the imprecise estimation of density for five different partition positions. This experiment

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

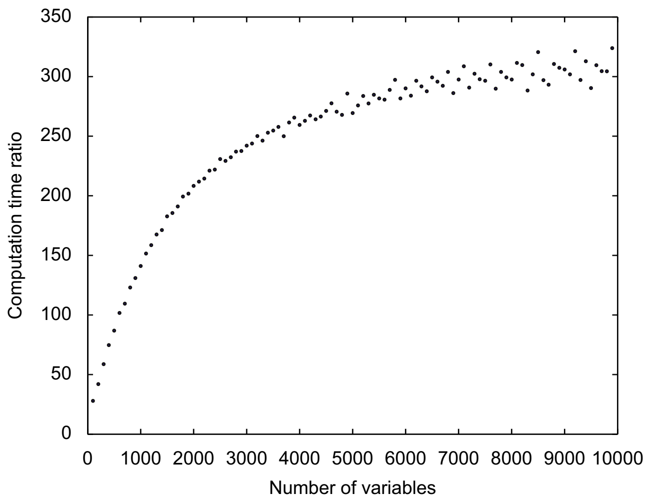


Fig. 8. Ratio of the computation time of the kernel density estimate to the computation time of the QCH density estimate.

illustrates the robustness of imprecise transfer for translation of the partition: all the interval-valued estimations have a non-empty intersection.

5.4. *Counting transfer and data precision*

When the observation precision is known, this information can be easily propagated on the estimated density. To illustrate this principle, we will assume that the precision of all observed eruption durations is equal to 0.025 min. Figs. 11–13 show estimates of duration distribution with kernels whose granularity equals 0.3 min. The QHC estimate is plotted in blue and red while usual kernel estimation (using the nine kernels defined previously) are plotted in cyan. Figs. 11a, 12a and 13a show the density estimate when the data are supposed to be precise, while Figs. 11b, 12b and 13b show the same experiment while accounting for imprecision. The imprecision of the estimated density increases when accounting for imprecision in observation. Thus, a known imprecision clearly impacts the precision of the density estimate. Moreover, when accounting for imprecision, the interval-valued estimation involving imprecise pignistic transfer is less specific. Therefore, as can be noted by comparing Figs. 12a and b, bracketing variations in conventional density estimation due to variation in the shapes of the involved summative kernels is better than in the precise case.

6. Conclusion

The technique we have presented in this paper consists of constructing a histogram of a set of observations on a fuzzy partition of the real line. Such histograms are able to account for precise and imprecise observations: a known imprecision in observation can easily be transferred to the histogram, which becomes bipolar, i.e. the counting is imprecise. Since the histogram is constructed on a fuzzy partition, it is possible to transfer the votes polling each cell of the partition to any bounded subset intersecting the partition, regardless of the sampled nature of the histogram. This counting transfer combined with the distributed vote is called the *quasi-continuous histogram* (QCH) *technique*. The ratio of the estimated number of observations purporting a given subset of the real line to the number of observations can be used to compute an objective confidence level associated with this subset.

We have proposed three kinds of counting transfer to perform this estimation. Precise pignistic transfer assumes that the density in each cell is uniform and that the cells are non-interactive. When accounting for cell interactivity, we shift to imprecise pignistic counting transfer. Then we shift to possibilistic transfer by freeing the transfer from any hypothesis on the distribution of votes on each cell. Imprecise pignistic transfer and possibilistic transfer partially propagate the ambiguity, induced by the distributed vote on the partition, on the inaccuracy of the counts transferred. Except when precise pignistic transfer with precise observation is used, the confidence level associated with a given

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

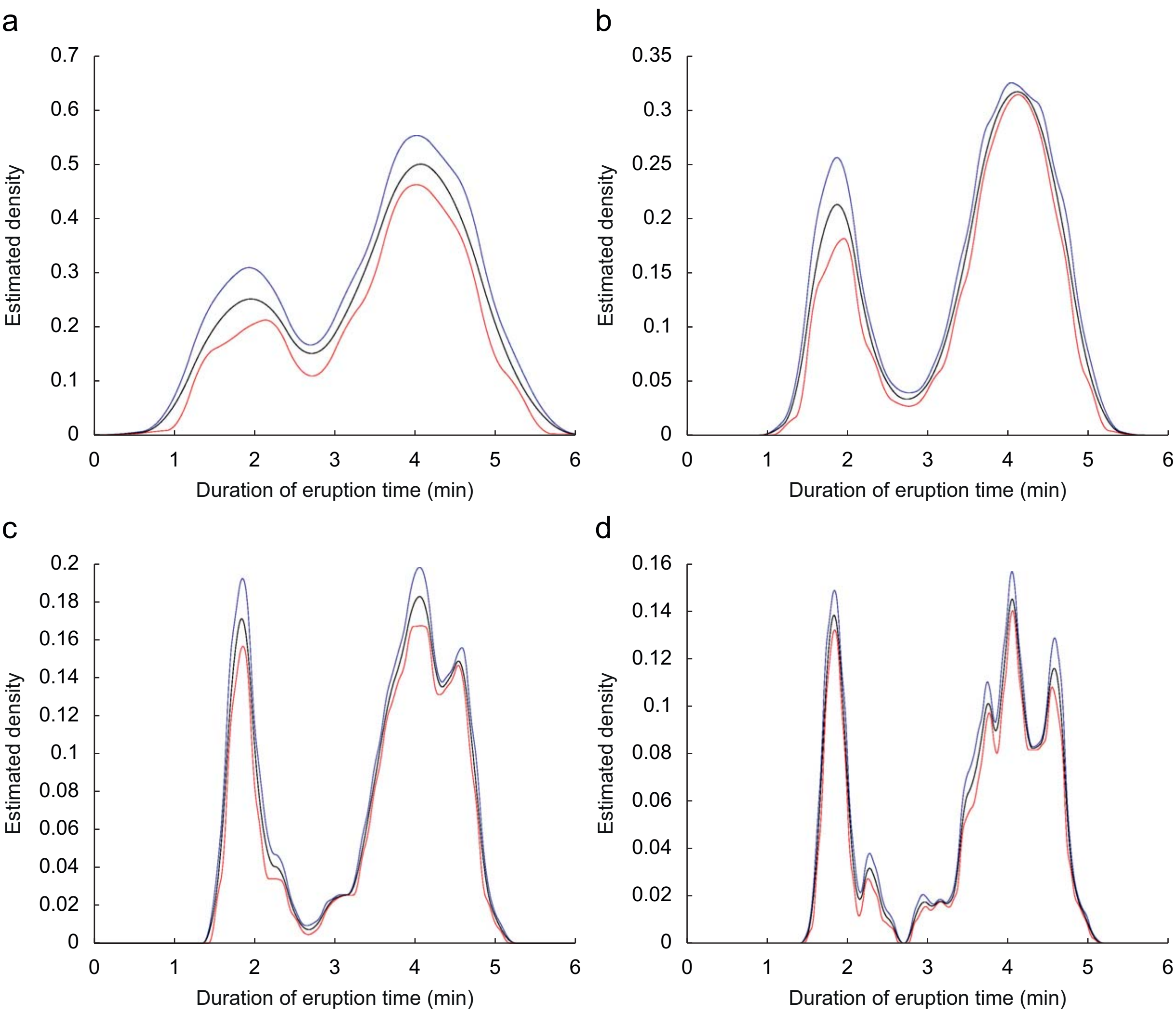


Fig. 9. Precise (black) and imprecise (blue—upper, red—lower) pignistic density estimation with different granularities ($a - \Delta = 0.60$, $b - \Delta = 0.30$, $c - \Delta = 0.15$, $d - \Delta = 0.10$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

subset is not a usual (additive) probability measure but rather an imprecise probability measure. Therefore, tools using this confidence measure to perform a given estimation may be adapted to account for this non-additivity.

The main advantage of using QCH is the possibility of obtaining a functional expression for precise or imprecise density probability estimation at a given location (for a given granularity). These functional expressions then allow the computation of statistical estimates based on the underlying putative density with algorithms whose computational complexity is proportional to the number of cells in the histogram. Such computations have been proposed in the past [2,5,25,32,41,44]. All of them involve what we have called pignistic counting transfers, except [41] for centile estimation. The validity of most of these tools still need to be mathematically proved. Compared to classic density estimation, QCH-based density estimation is fast, accounts for a known degree of precision in observation, and gives an estimate of the impact of the partitioning on the estimated density. The choice of kernel shape and bandwidth is still arbitrary, but this arbitrariness is propagated on the imprecision of the density estimation, when an imprecise counting transfer is chosen. We think, however, that the best choice for partitioning the fpbh is the triangular kernel. This opinion is partly, but not solely, motivated by the specific role of this kernel in the possibilistic measure context. In fact, due to Property 7, the QCH-based density estimation (with precise pignistic transfer) is equivalent to a Parzen Rosenblatt

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013

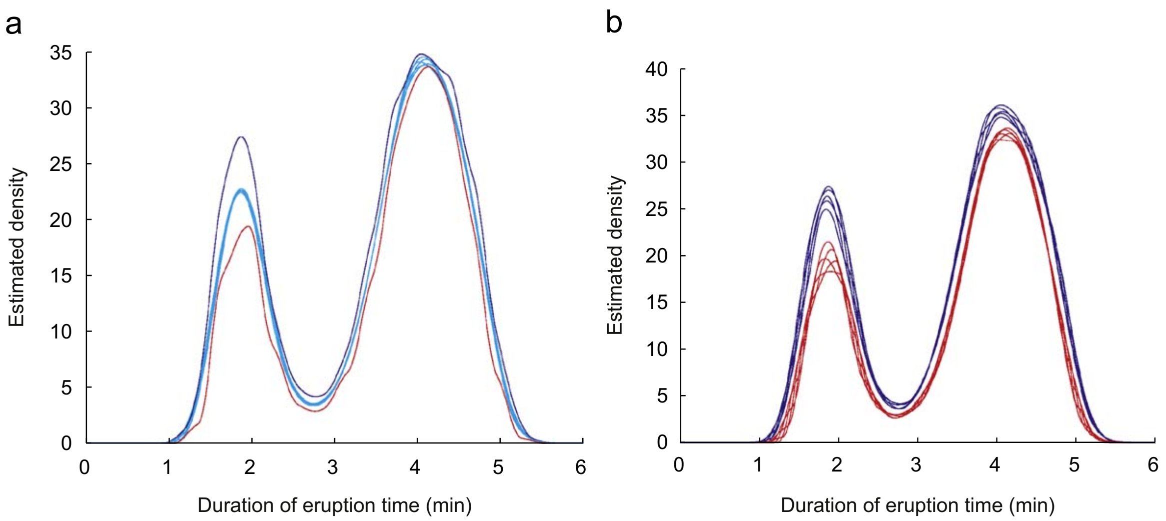


Fig. 10. Translation of the fuzzy partition, (a) five superimposed precise estimations with one imprecise estimation, (b) five superimposed imprecise estimations.

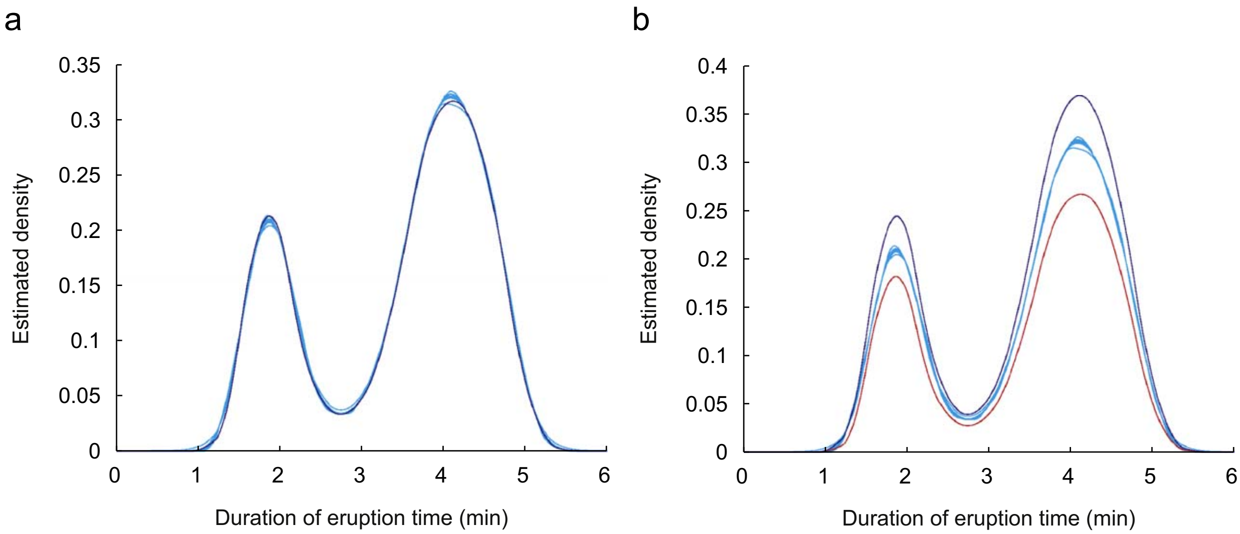


Fig. 11. Precise pignistic transfer with precise (a) and imprecise (b) observations.

density estimation based on a kernel whose shape is given by convolving the basic kernel of the partition with itself. Convolving two triangular kernels provides an Epanechnikov kernel, which is usually considered to be the optimal choice in a density estimation context, since it minimizes the asymptotic mean integrated square error performance criterion. Moreover, the use of a triangular kernel-based partition leads to very simple computation.

The counting transfers we have proposed are based on generalizing crisp counting transfer principles by using a Choquet integral. This generalization aims at providing simple expressions for computing interval-valued densities. Other generalizations can be obtained for example by considering the Sugeno integral or by constructing a fuzzy valued density as in [45,39]. Such a construction could be envisaged by transferring the counts to a fuzzy subset instead of a crisp subset of Ω. Moreover, different ways of constructing histograms could be considered involving voting not only

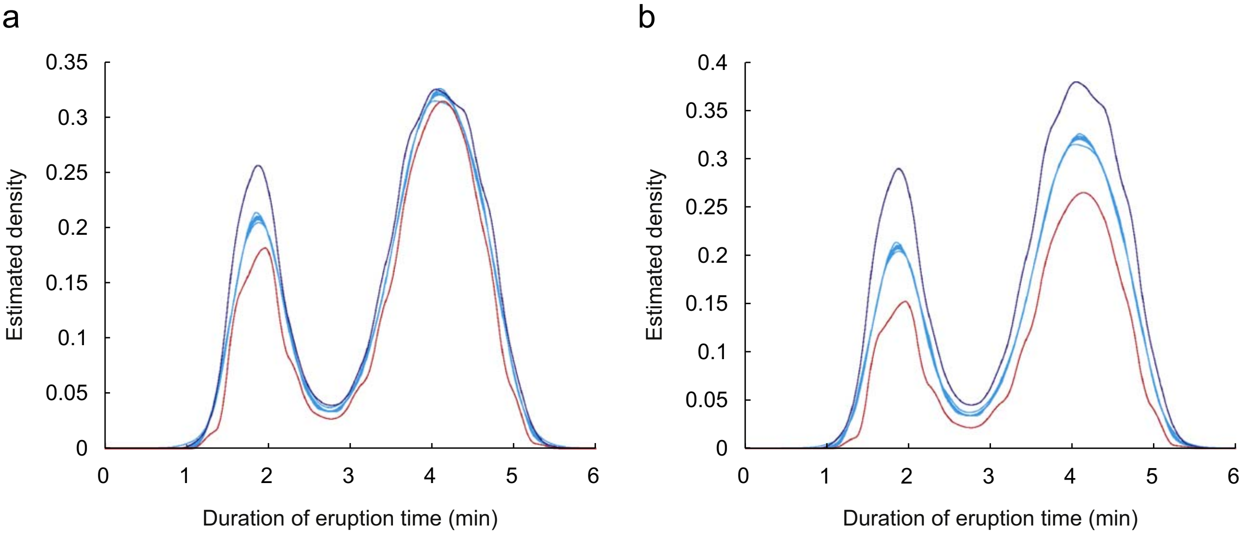


Fig. 12. Imprecise pignistic transfer with precise (a) and imprecise (b) observations.

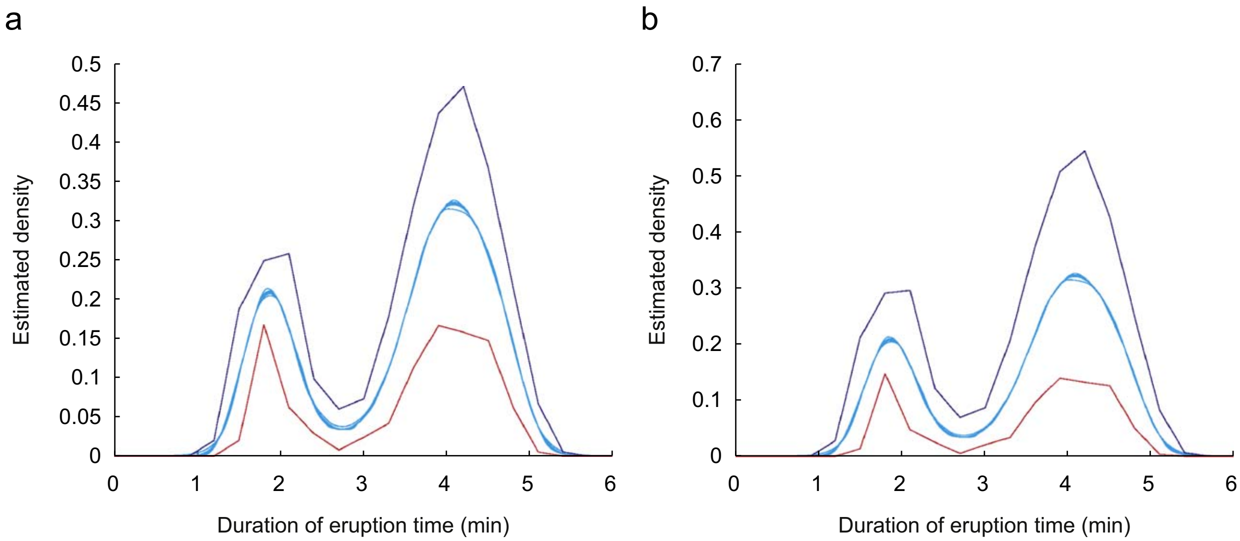


Fig. 13. Possibilistic transfer with precise (a) and imprecise (b) observations.

for each cell of the partition but also for any union of cells in a more random-set way. This kind of voting would be more likely to properly handle imprecise observations by avoiding the risk of weak accumulations.

References

[1] G. Derzko, Une approche intrinsèque de l'estimation non paramétrique de la densité, Comptes Rendus de l'Académie des Sciences–Série I–Mathematics 327 (1998) 985–988.
[2] H. Cheng, Y. Chen, Fuzzy partition of two-dimensional histogram and its application to thresholding, Pattern Recognition 32 (1999) 825–843.
[3] V. Chatzis, I. Pitas, Fuzzy cell Hough transform for curve detection, Pattern Recognition 30 (1997) 2031–2042.
[4] L. Chmielewski, Fuzzy histograms, weak fuzzification and accumulation of periodic quantities, Pattern Analysis and Applications 9 (2006) 189–210.
[5] F. Comby, O. Strauss, Using quasi-continuous histograms for fuzzy main motion estimation in video sequence, Fuzzy Sets and Systems 158 (2007) 475–495.
[6] G. de Cooman, D. Aeyels, Supremum preserving upper probabilities, Information Sciences 118 (1998) 173–212.

[7] A. Dempster, Upper and lower probabilities induced by a multivalued mapping, The Annals of Mathematical Statistics 38 (1967) 325–339.
[8] D. Denneberg, Non Additive Measure and Integral, Kluwer Academic Publishers, Dordrecht, 1994.
[9] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–200.
[10] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, London, 1985.
[11] D. Dubois, H. Prade, S. Sandri, On possibility/probability transformations, in: R. Lowen, M. Roubens (Eds.), Fuzzy Logic: State of the Art, Kluwer Academic Publishers, Dordrecht, 1993, pp. 103–112.
[12] D. Dubois, L. Foulloy, G. Mauris, H. Prade, Probability-possibility transformation, triangular fuzzy sets, and probabilistic inequalities, Reliable Computing 10 (2004) 273–297.
[13] D. Dubois, Possibility theory and statistical reasoning, Computational Statistics and Data Analysis 51 (2006) 47–69.
[14] M. Grabisch, T. Murofushi, M. Sugeno (Eds.), Fuzzy Measures and Integrals: Theory and Applications, Physica, Berlin, 2000.
[15] J. Han, L. Kóczy, T. Poston, Fuzzy Hough transform, Pattern Recognition Letters 15 (1994) 649–658.
[16] T. Hastie, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, Berlin, 2003.
[17] L. Jaulin, M. Kieffer, O. Didrit, Walter E, Applied Interval Analysis, Springer, Berlin, 2001.
[18] L. Jaulin, Computing minimal-volume credible sets using interval analysis; application to Bayesian estimation, IEEE Transactions on Signal Processing 54 (2006) 3632–3636.
[19] K. Loquin, O. Strauss, On the granularity of summative kernels, Fuzzy Sets and Systems 159 (2008) 1952–1972.
[20] K. Loquin, O. Strauss, Fuzzy histograms and density estimation, in: SMPS 2006, Third Internat. Workshop on Soft Methods in Probability and Statistics, September 5–7, 2006, pp. 45–52.
[21] K. Loquin, O. Strauss, Histogram density estimators based upon a fuzzy partition, Statistics and Probability Letters 78 (2008) 1863–1868.
[22] J. Marichal, An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria, IEEE Transactions on Fuzzy Systems 8 (2000) 800–807.
[23] E. Miranda, I. Couso, P. Gil, Random intervals as a model for imprecise information, Fuzzy Sets and Systems 154 (2005) 386–412.
[24] P. Matsakis, S. Andréfouët, P. Capolsini, Evaluation of fuzzy partitions, Remote Sensing of Environment 74 (2000) 516–533.
[25] P. Matsakis, L. Wendling, A new way to represent the relative position between areal objects, IEEE Transactions on PAMI 21 (1999) 634–643.
[26] J. Nuñez-Garcia, O. Wolkenhauer, Random sets and histograms, in: the 10th IEEE Internat. Conf. on Fuzzy Systems, vol. 3, Melbourne, Australia, 2001, pp. 1183–1186.
[27] E. Parzen, On estimation of a probability density function and mode, Annals of Mathematical Statistics 33 (1962) 1065–1076.
[28] E. Parzen, Non parametric statistical data modeling, Journal of American Statistical Association 74 (1979) 105–131.
[29] I. Perfilieva, Fuzzy function as an approximate solution to a system of fuzzy relation equations, Fuzzy Sets and Systems 147 (2004) 363–383.
[30] A. Ralescu, D. Ralescu, Extensions of fuzzy aggregation, Fuzzy Sets and Systems 86 (1997) 321–330.
[31] S. Ralescu, The law of the iterated logarithm for the multivariate nearest neighbor density estimators, Journal of Multivariate Analysis 53 (1995) 159–179.
[32] T. Runkler, Fuzzy histograms and fuzzy chi-squared tests for independance, in: FUZZ-IEEE 2004, IEEE Internat. Conf. on Fuzzy Systems, vol. 3, 2004, pp. 1361–1366.
[33] E. Ruspini, New experimental results in fuzzy clustering, Information Sciences 6 (1973) 273–284.
[34] D. Schmeidler, Integral representation without additivity, Proceedings of the American Mathematical Society 97 (1986) 255–261.
[35] D. Scott, S. Sheather, Kernel density estimation with binned data, Communications in Statistics, Theory and Methods 14 (1985) 1353–1359.
[36] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, CRC, London, 1986.
[37] J. Simonoff, Smoothing Methods in Statistics, Springer, Berlin, Germany, 1996.
[38] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, International Journal of Approximate Reasoning 38 (2005) 133–147.
[39] W. Trutschnig, A strong consistency result for fuzzy relative frequencies interpreted as estimator for the fuzzy-valued probability, Fuzzy Sets and Systems 159 (2008) 259–269.
[40] O. Strauss, Use the fuzzy Hough transform (towards reduction of the precision/uncertainty duality), Pattern Recognition 32 (1999) 1911–1922.
[41] O. Strauss, F. Comby, M.J. Aldon, Rough histograms for robust statistics, in: ICPR'2000 15th Internat. Conf. on Pattern Recognition, Barcelona, Catalonia, Spain, 2000, pp. 688–691.
[42] N. Suetake, E. Uchino, K. Hirata, Generalized fuzzy Hough transform for detecting arbitrary shapes in a vague and noisy image, Soft Computing 10 (2006) 1161–1168.
[43] W. van den Bergh, J. van den Berg, Competitive exception learning using fuzzy frequency distributions, Neural Network World 10 (2000) 59–71.
[44] J. Van den Berg, W. Van den Bergh, U. Kaymak, Probabilistic and statistical fuzzy set foundations of competitive exception learning, in: 10th IEEE Internat. Conf. on Fuzzy Systems, vol. 2, Melbourne, Australia, 2001, pp. 1035–1038.
[45] R. Viertl, W. Trutschnig, Fuzzy histograms and fuzzy probability distributions, in: 11th Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems, Editions EDK, 2006.
[46] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman & Hall, New York, 1991.
[47] L. Waltman, U. Kaymak, J. van den Berg, Fuzzy histograms: a statistical analysis, in: 4th Conf. of the European Society for Fuzzy Logic and Technology, Barcelona, Spain, 2005, pp. 605–610.

Please cite this article as: O. Strauss, Quasi-continuous histograms, Fuzzy Sets and Systems (2009), doi:10.1016/j.fss.2009.01.013