
Fuzzy histograms and density estimation

LOQUIN Kevin

LIRMM - 161 rue Ada
34392 Montpellier cedex 5 - France
loquin@lirmm.fr

The *probability density function* is a fundamental concept in statistics. Specifying the density function f of a random variable X on Ω gives a natural description of the distribution of X on Ω . When it cannot be specified, an estimate of this density may be performed by using a sample of n observations iid (X_1, \dots, X_n) of X .

Histogram is the oldest and most widely used density estimator for presentation and exploration of observed univariate data. The construction of a histogram consists in partitioning a given reference interval Ω into p bins A_k and to count the number Acc_k of observations that belong to each cell A_k . If all the A_k have the same width h , the histogram is said to be uniform or regular. Let $\mathbb{1}_{A_k}$ be the characteristic function of A_k , we have

$$Acc_k = \sum_{i=1}^n \mathbb{1}_{A_k}(X_i) \quad (1)$$

By hypothesizing a uniform density of the data observed in each cell, an estimate $\hat{f}_{hist}(x)$ of the underlying probability density function $f(x)$ at any point x of A_k can be computed by:

$$\hat{f}_{hist}(x) = \frac{Acc_k}{nh} \quad (2)$$

Its popularity is due not only to its simplicity (no particular skills needed to manipulate this tool) but also to the fact that the piece of information provided by a histogram is more than a rough representation of the density underlying the data. In fact, a histogram displays the number of data (or observations) of a finite data set that belong to a given class i.e. in complete agreement with the concept summarized by the label associated with each bin of the partition thanks to the quantity Acc_k .

However, the histogram density estimator has some weaknesses. The approximation given by expression (2) is a discontinuous function. The choice of both reference interval and number of cells (i.e. bin width) have quite an

effect on the estimated density. The apriorism needed to set those values make it a tool whose robustness and reliability are too low to be used for statistical estimation.

In the last five years, it has been suggested by some authors to low the effect of arbitrariness of partitioning by replacing the binary partition by a fuzzy partition. This solution has been studied as a practical tool for Chi-squared tests [Run04], estimation of conditional probabilities in a learning context [VDB01], or estimation of percentiles [SCA00] and modes [SC02]. Fuzzy partitioning has received considerable attention in the literature especially in the field of control and decision theory. Recently, some authors have proposed to explore the universal approximation properties of fuzzy systems to solve system of equations [Per, Per04a, Per04b, Wan98, HKAS03, HKAS06, Lee02].

In a first part, we will formally present the fuzzy partition as proposed in [Per]. Secondly, an histogram based upon this previous notion will be defined, we will call it a fuzzy histogram. And in a last section some estimators of probability density functions will be shown, before concluding.

1 Fuzzy partitions

1.1 Preliminary

Prior to formally defining a fuzzy partition, let me remind you that the Fuzzy Set Theory is nothing else than a softened generalization of the Classical Set Theory. Indeed, in the latter, the decision whether the element x belongs to a subset F of Ω , the universe, is tantamount to the question whether it is true that $x \in F$ (this is a binary question). However, in many practical cases, this question cannot be precisely answered : there exists a vagueness in the "frontiers" of F . A reasonable solution consists in using a scale whose elements would express various *degrees of truth* of $x \in F$. Let L be this scale of truth values. We usually put $L = [0, 1]$. Then

Definition 1. *The fuzzy subset F is identified with its membership function*

$$\mu_F : \Omega \rightarrow L = [0, 1]$$

assigning the value $\mu_F(x) \in L$ to each element $x \in \Omega$ which is the membership degree of x in F . We will note $F \sqsubset \Omega$ if F is a fuzzy subset of Ω . The set of all fuzzy subsets of Ω is $\mathfrak{F}(\Omega) = \{F \mid F \sqsubset \Omega\} = L^\Omega$. Note that $\mathfrak{F}(\Omega)$ contains the classical subsets of Ω . Indeed, if F is an classical subset of Ω , we take $\mu_F(x) = \mathbb{1}_F(x)$.

Once recalled this basic notion, We will now replace the classical subsets which form a partition of the universe, by fuzzy subsets.

1.2 Strong Uniform Fuzzy Partition of the Universe

Here we will take an interval $\Omega = [a, b]$ (real) as the universe. Then,

Definition 2. Let $m_1 < m_2 < \dots < m_p$ be p fixed nodes of the universe, such that $m_1 = a$ and $m_p = b$, and $p \geq 3$. We say that the set of the p fuzzy subsets A_1, A_2, \dots, A_p , identified with their membership functions $\mu_{A_1}(x), \mu_{A_2}(x), \dots, \mu_{A_p}(x)$ defined on the universe, form a strong uniform fuzzy partition of the universe, if they fulfil the following conditions :

for $k = 1, \dots, p$

1. $\mu_{A_k}(m_k) = 1$ (m_k belongs to what is called the core of A_k),
2. if $x \notin [m_{k-1}, m_{k+1}]$, $\mu_{A_k}(x) = 0$ (because of the notation we should add : $m_0 = m_1 = a$ and $m_p = m_{p+1} = b$),
3. $\mu_{A_k}(x)$ is continuous,
4. $\mu_{A_k}(x)$ monotonically increases on $[m_{k-1}, m_k]$ and $\mu_{A_k}(x)$ monotonically decreases on $[m_k, m_{k+1}]$,
5. $\forall x \in \Omega, \exists k$, such that $\mu_{A_k}(x) > 0$ (every element of the universe is treated in this partition).
6. for all $x \in \Omega, \sum_{k=1}^p \mu_{A_k}(x) = 1$
7. for $k \neq p, h_k = m_{k+1} - m_k = h = \text{constant}$, so, $m_k = a + (k-1)h$,
8. for $k \neq 2$ and $k \neq p, \forall x \in [0, h] \mu_{A_k}(m_k - x) = \mu_{A_k}(m_k + x)$ (μ_{A_k} is symmetric around m_k),
9. for $k \neq 2$ and $k \neq p, \forall x \in [m_k, m_{k+1}], \mu_{A_k}(x) = \mu_{A_{k-1}}(x - h)$ and $\mu_{A_{k+1}}(x) = \mu_{A_k}(x - h)$ (all the μ_{A_k} , for $k = 2, \dots, p-1$ have the same shape, with a translation of h . And as for μ_{A_1} and μ_{A_p} , they have the same shape, but truncated, with supports twice smaller than the other ones).

We can add that the condition 6. is the strength condition, i.e. that without this condition the fuzzy partition is no more strong. Moreover, we can stress the importance of this condition, because it gives a *normal weight* of 1, to each element x of the universe in the partition. In the same way, the conditions 7., 8. and 9. are the uniformity conditions.

And now, comes a quasi-immediate proposition

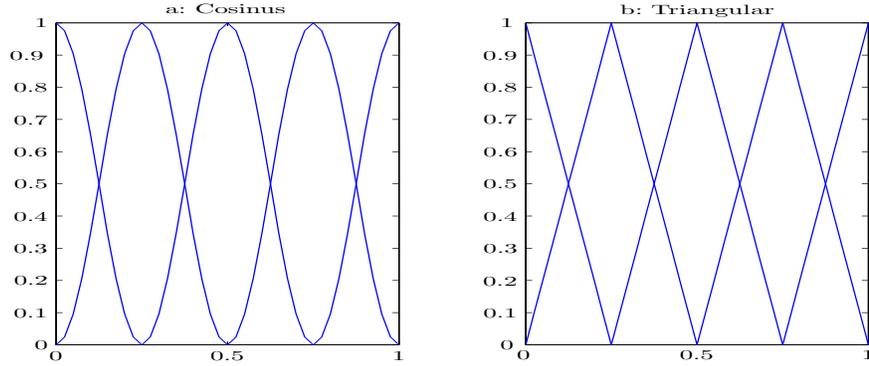
Proposition 1. Let $(A_k)_{k=1, \dots, p}$ be a strong uniform fuzzy partition of the universe, then

$\exists K_A : [-1, 1] \rightarrow [0, 1]$ pair, such that, $\mu_{A_k}(x) = K_A(\frac{x-m_k}{h}) \mathbb{1}_{[m_{k-1}, m_{k+1}]}$ and $\int K_A(u) du = 1$.

Proof. We can take $K_A(u) = \mu_{A_k}(hu + m_k), \forall k$. The support of K_A comes from the ones of the μ_{A_k} , and the parity, is deduced from a translation of the symmetry of the μ_{A_k} . And eventually, $\int_{-1}^1 K_A(u) du = \int_{-1}^1 \mu_{A_k}(hu + m_k) du = \int_{m_{k-1}}^{m_{k+1}} \frac{1}{h} \mu_{A_k}(x) dx = 1$.

Table 1. Strong uniform fuzzy partition examples

	Crisp	Triangular	Cosinus
$\mu_{A_1}(x) =$	$\mathbb{1}_{[m_1, m_1 + \frac{h}{2}]}(x)$	$\frac{(m_2 - x)}{h} \mathbb{1}_{[m_1, m_2]}(x)$	$\frac{1}{2} (\cos(\frac{\pi(x - m_1)}{h}) + 1) \mathbb{1}_{[m_1, m_2]}(x)$
$\mu_{A_k}(x) =$	$\mathbb{1}_{[m_k - \frac{h}{2}, m_k + \frac{h}{2}]}(x)$	$\frac{(x - m_{k-1})}{h} \mathbb{1}_{[m_{k-1}, m_k]}(x)$ + $\frac{(m_{k+1} - x)}{h} \mathbb{1}_{[m_k, m_{k+1}]}(x)$	$\frac{1}{2} (\cos(\frac{\pi(x - m_k)}{h}) + 1) \mathbb{1}_{[m_{k-1}, m_{k+1}]}(x)$
$\mu_{A_p}(x) =$	$\mathbb{1}_{[m_p - \frac{h}{2}, m_p]}(x)$	$\frac{(x - m_{p-1})}{h} \mathbb{1}_{[m_{p-1}, m_p]}(x)$	$\frac{1}{2} (\cos(\frac{\pi(x - m_p)}{h}) + 1) \mathbb{1}_{[m_{p-1}, m_p]}(x)$
$K_A(x) =$	$\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(x)$	$(1 - x) \mathbb{1}_{[-1, 1]}(x)$	$0.5 (\cos(\pi x) + 1) \mathbb{1}_{[-1, 1]}(x)$

**Fig. 1.** Fuzzy partitions with $\Omega = [0, 1]$ and $p = 5$

2 A fuzzy-partition based histogram

The accumulated value Acc_k is the key feature of the histogram technique. It is the number of observations in complete agreement with the label represented by the restriction of the real line to the interval (or bin) A_k . Due to the important arbitrariness of the partition, the histogram technique is known as being very sensitive to the choice of both reference interval and number of cells (or bin width). As mentioned before, the effect of this arbitrariness can be reduced by replacing the crisp partition by a fuzzy partition of the real line.

Let $(A_k)_{k=1, \dots, p}$ be a strong uniform fuzzy partition of Ω , the natural extension of the expression (1) induces a distributed vote. The value of the accumulator Acc_k associated to the fuzzy subset A_k is given by:

$$Acc_k = \sum_{i=1}^n \mu_{A_k}(X_i) \quad (3)$$

Then, those accumulators still represent a "real" (generally not an integer) number of observations in accordance with the label represented by the fuzzy

subset A_k . Moreover, the *strength*¹ of the fuzzy partition $(A_k)_{k=1,\dots,p}$ implies that the sum of the Acc_k equals to n , the number of observations. Having said that, we can remark that the classical crisp-partition based histogram is a particular case of the fuzzy-partition based histogram, when $(A_k)_{k=1,\dots,p}$ is the crisp partition.

We propose, by mean of the figures 2 an illustration of the softening property of the fuzzy histogram over the the crisp histogram. (a) displays a crisp histogram of 35 observations drawn from a Gaussian process with mean $\mu = 0.3$ and variance $\sigma^2 = 1$. (c) displays a fuzzy-partition based histogram of the same observations with the same reference interval position. We have translated the latter of 30% of the bin width. It can be seen on (b) that this translation had quite an effect on the crisp-partition based histogram, while the fuzzy-partition based histogram plotted on the (d) still has the same general shape.

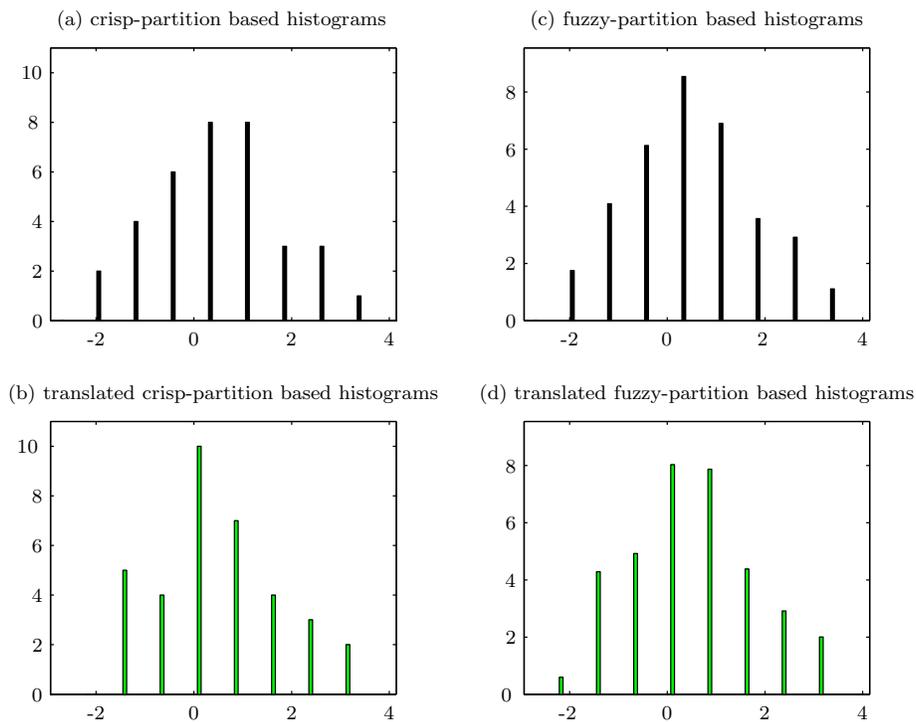


Fig. 2. Partition translation effect on crisp ((a) and (b)) and fuzzy ((c) and (d)) histograms

¹ Every observation X_i counts for 1 in the fuzzy partition, as the condition 6. of the definition 2 suggests.

3 Fuzzy histogram density estimators

Afterwards, we can, as in the crisp histogram case, deduce from this histogram, by supposing a uniform repartition of the data around each bin, an estimate $\hat{f}_{FH}(x)$ of the underlying probability density function $f(x)$ at any point x of $[m_k - h/2, m_k + h/2]$, which can be computed by:

$$\hat{f}_{FH}(x) = \frac{Acc_k}{nh} \quad (4)$$

This fuzzy histogram density estimator is better than the crisp histogram density estimator of the expression (2) in the sense that it is more robust to partition translations or dilations. However, its discontinuity is kept. Then, in order to cope with this flaw, we propose a different approach to construct density estimators from fuzzy (or crisp) histograms.

First, we must remark that a fuzzy histogram is actually a set of p pairs (A_k, Acc_k) or p points (m_k, Acc_k) , when reducing each fuzzy subset A_k to its node m_k . Then, assuming that our density estimators have to go through the p points $(m_k, \frac{Acc_k}{nh})$, an interpolation (see [Kop55]) appears naturally as an adequate solution.

In this paper we propose to use once again the concept of fuzzy partition, to interpolate the normalized histogram (represented by the p points above given). In other words,

Proposition 2. *An interpolant of a fuzzy histogram (of the points $(m_k, \frac{Acc_k}{nh})$) is given by*

$$\hat{f}_{FH}(x) = \frac{1}{nh} \sum_{k=1}^p Acc_k \mu_{B_k}(x) \quad (5)$$

where $(B_k)_{k=1, \dots, p}$ is a strong uniform fuzzy partition.

Proof. The conditions 1 and 6 of the definition 2 imply that $\mu_{B_k}(m_l) = \delta_{kl}$ for $k, l \in \{1, \dots, p\}$, where δ_{kl} is the Kronecker symbol. Then, $\hat{f}_{FH}(m_l) = \frac{Acc_l}{nh}$, for all $l \in \{1, \dots, p\}$, which means that \hat{f}_{FH} goes through the p points $(m_k, \frac{Acc_k}{nh})$.

Therefore, this interpolant (which is a density estimator) has the continuity properties of the membership functions of the fuzzy partition $(B_k)_{k=1, \dots, p}$, except to the nodes m_k , where the smoothness is not guaranteed. We can now add a convergence property of the estimators (5). So, let us take as measure of error between the underlying density $f(x)$ and the estimate $\hat{f}_{FH}(x)$, the mean squared error (MSE), defined by $MSE(x) \triangleq E_f[\hat{f}_{FH}(x) - f(x)]^2$. We have proved in a paper in preparation, the following theorem.

Theorem 1. *Let us suppose*

1. $f : \Omega \rightarrow [0, 1]$ is a density function such that f is bounded ($\forall x \in \Omega$, $f(x) \leq f_{max} < +\infty$) and f' is bounded ($\forall x \in \Omega$, $|f'(x)| \leq f'_{max} < +\infty$),

2. K , as defined in proposition 1, verifies $\int_{-1}^1 K^2(u)du < +\infty$.

Then, for all $x \in \Omega$,

$$h \rightarrow 0 \text{ and } nh \rightarrow +\infty \Rightarrow MSE(x) \rightarrow 0 \tag{6}$$

This theorem gives us mathematical evidence that the fuzzy histogram is a proper representation of the distribution of data, because a simple interpolation of a normalized histogram converges (in MSE) to the underlying density. It converges under classical conditions, which are, the reduction of the support of the membership functions, or the growth of the number of fuzzy subsets of the partition ($h \rightarrow 0$), and the growth of the mean number of data in each accumulator ($nh \rightarrow +\infty$).

However, the use of the membership functions of a fuzzy partition as interpolation functions is not compulsory. And well known interpolation functions could be used instead of them. We can here enumerate the polynomial interpolation (with the Lagrange form, see the yellow plotted curves of the figure 3, or the Newton form), the interpolation by rational functions, the trigonometric interpolation, and last but not least the the spline interpolation, which has the advantage to manage the smoothness at the nodes, see the cyan plotted curves of the figure 3.

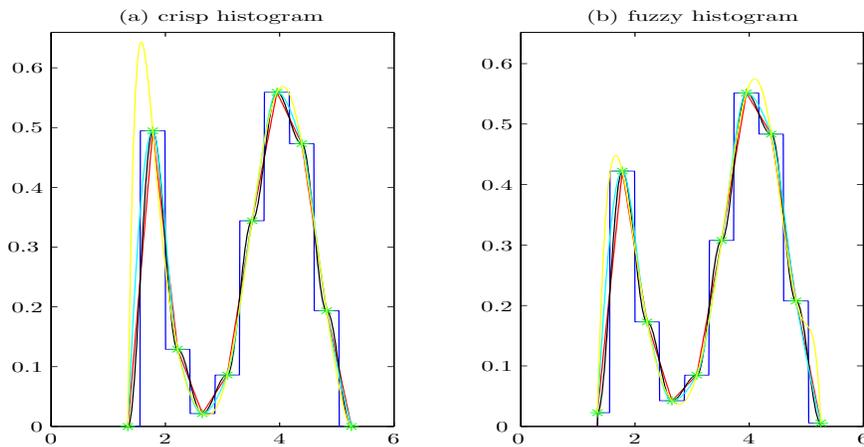


Fig. 3. Density estimation by interpolation

In this previous figure, we can observe estimations of the distribution of eruption lengths of *Old Faithful Geyser* at Yellowstone (USA), based upon a crisp histogram, for (a) an a fuzzy (triangular in that case) histogram for (b). The green stars represent the histograms, the blue curves are the crisp

estimation (see expression (4)), the red and black ones are the estimators by fuzzy histogram interpolation (see expression (5)) with respectively triangular and cosine membership functions.

References

- [HKAS03] R. Hassine, F. Karray, A. M. Alimi, and M. Selmi. Approximation properties of fuzzy systems for smooth functions and their first order derivative. *IEEE Trans. Systems Man Cybernet.*, 33:160–168, 2003.
- [HKAS06] R. Hassine, F. Karray, A. M. Alimi, and M. Selmi. Approximation properties of piece-wise parabolic functions fuzzy logic systems. *Fuzzy sets and systems*, 157:501–515, 2006.
- [Kop55] Z. Kopal. *Numerical analysis*. Wiley New York, 1955.
- [Lee02] T. Leephakpreeda. Novel determination of differential-equation solutions : universal approximation method. *Journal of computational and applied mathematics*, 146:443–457, 2002.
- [Per] I. Perfilieva. Fuzzy transforms. *Institute for Research and Applications of Fuzzy Modelling, University of Ostrava*.
- [Per04a] I. Perfilieva. Fuzzy function as an approximate solution to a system of fuzzy relation equations. *Fuzzy sets and systems*, 147:363–383, 2004.
- [Per04b] I. Perfilieva. Normal forms in bl-algebra and their contribution to universal approximation of functions. *Fuzzy sets and systems*, 143:111–127, 2004.
- [Run04] T. A. Runkler. Fuzzy histograms and fuzzy chi-squared tests for independence. *IEEE international conference on fuzzy systems*, 3:1361–1366, 2004.
- [SC02] O. Strauss and F. Comby. Estimation modale par histogramme quasi-continu. *LFA'02 Rencontres Francophones sur la Logique Floue et ses Applications, Montpellier*, pages 35–42, 2002.
- [SCA00] O. Strauss, F. Comby, and M.J. Aldon. Rough histograms for robust statistics. *ICPR'2000 15th International Conference on Pattern Recognition, Barcelona, Catalonia, Spain, 3-8 September, 2000*.
- [VDB01] J. Van Den Berg. Probabilistic and statistical fuzzy set foundations of competitive exception learning. *IEEE International Fuzzy Systems Conference*, pages 1035–1038, 2001.
- [Wan98] L. X. Wang. Universal approximation by hierarchical fuzzy systems. *Fuzzy sets and systems*, 93:223–230, 1998.