Fuzzy Histograms And Density Estimation

Kevin Loquin, Olivier Strauss

Fuzzy histograms and density estimation

Kevin LOQUIN[1] and Olivier STRAUSS[2]

LIRMM - 161 rue Ada - 34392 Montpellier cedex 5 - France
[1]Kevin.Loquin@lirmm.fr
[2]Olivier.Strauss@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 the universe Ω. When it cannot be specified, an estimate of this density may be performed by using a sample of n observations independent and identically distributed $(X_1, ..., X_n)$ 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 in counting the number Acc_k of observations belonging 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

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

By hypothesizing the density of the data observed in each cell to be uniform, 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:

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

The popularity of the histogram technique is not only due to its simplicity (no particular skills are 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 makes 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 that replacing the binary partition by a fuzzy partition will reduce the effect of arbitrariness of partitioning. 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 [Per06, Per04, Wan98, HKAS03, Lee02].

In a first part, we will formally present the fuzzy partition as proposed in [Per06]. In section 2, a histogram based upon this previous notion will be defined, that will be called a fuzzy histogram. In a last section, some estimators of probability density functions will be shown, before concluding.

1 Fuzzy partitions

1.1 Preliminary

In histogram technique, the accumulation process (see expression (1)) is linked to the ability to decide whether the element x belongs to a subset A_k of Ω, the universe, or not. This decision is tantamount to the question whether it is true that $x \in A_k$ or not (this is a binary question). However, in many practical cases, this question cannot be precisely answered : there exists a vagueness in the "frontiers" of A_k. A reasonable solution consists in using a scale whose elements would express various *degrees of truth* of $x \in A_k$, and A_k becomes a fuzzy subset of Ω. Let L be this scale of truth values. We usually put $L = [0, 1]$.

1.2 Strong Uniform Fuzzy Partition of the Universe

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

Definition 1. *Let $m_1 < m_2 < ... < m_p$ 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,...,A_p, identified with their membership functions $\mu_{A_1}(x)$,$\mu_{A_2}(x)$,...,$\mu_{A_p}(x)$ defined on the universe, form a* strong uniform fuzzy partition of the universe*, if they fulfil the following conditions :*

for $k = 1, ..., p$

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 $k \neq p$, $h_k = m_{k+1} - m_k = h = constant$, so, $m_k = a + (k-1)h$,*
8. *for $k \neq 1$ 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 $k \neq 1$ 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 $k = 2, ..., p-1$ 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).*

Condition 6 is known as the strength condition, which ensures a *normal weight* of 1, to each element x of the universe in a strong fuzzy partition. In the same way, conditions 7, 8 and 9 are the conditions for the uniformity of a fuzzy partition.

Proposition 1. *Let $(A_k)_{k=1,...,p}$ be a strong uniform fuzzy partition of the universe, then*
$\exists K_A : [-1, 1] \longrightarrow [0, 1]$ 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, to end this proof, $\int_{-1}^{1} K_A(u)du = \int_{-1}^{1} \mu_{A_k}(hu + m_k)du = \int_{m_{k-1}}^{m_{k+1}} \frac{1}{h}\mu_{A_k}(x)dx = 1$.

Table 1. Strong uniform fuzzy partition examples

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

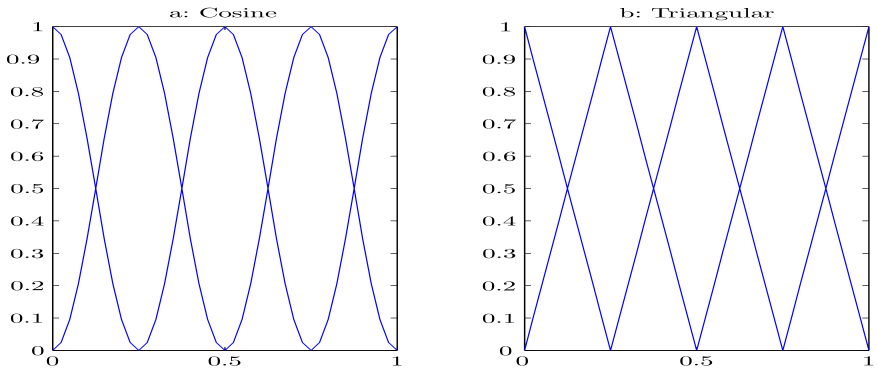

Fig. 1. Fuzzy partitions with $\Omega = [0, 1]$ and $p = 5$

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 $(A_k)_{k=1,\ldots,p}$ 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:

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

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* (Condition 6 of Definition 1) of the fuzzy partition $(A_k)_{k=1,\ldots,p}$ implies that the sum of the Acc_k equals to n,[1] the number of observations. Note that the classical crisp-partition based histogram is a particular case of the fuzzy-partition based histogram, when $(A_k)_{k=1,\ldots,p}$ is the crisp partition.

We propose to illustrate the softening property of the fuzzy histogram over the crisp histogram. Figure 2.(a) displays a crisp histogram of 35 observations drawn from a Gaussian process with mean $\mu = 0.3$ and variance $\sigma^2 = 1$. Figure 2.(b) displays a fuzzy triangular partition based histogram of the same

[1] indeed, $\sum_{k=1}^{p} Acc_k = \sum_{k=1}^{p} \sum_{i=1}^{n} \mu_{A_k}(X_i) = \sum_{i=1}^{n} \sum_{k=1}^{p} \mu_{A_k}(X_i) \sum_{i=1}^{n} 1 = n$

observations with the same reference interval position. We have translated both crisp and fuzzy partitions by an amount of 30% of the bin width. As it can be seen on Figure 2.(c), this translation has quite an effect on the crisp-partition based histogram, while the fuzzy-partition based histogram plotted on Figure 2.(d) still has the same general shape. The number of observations is too small, regarding the number of fuzzy subsets ($p = 8$) of the partition, to ensure that the convergence conditions are fulfilled (see theorem 1).

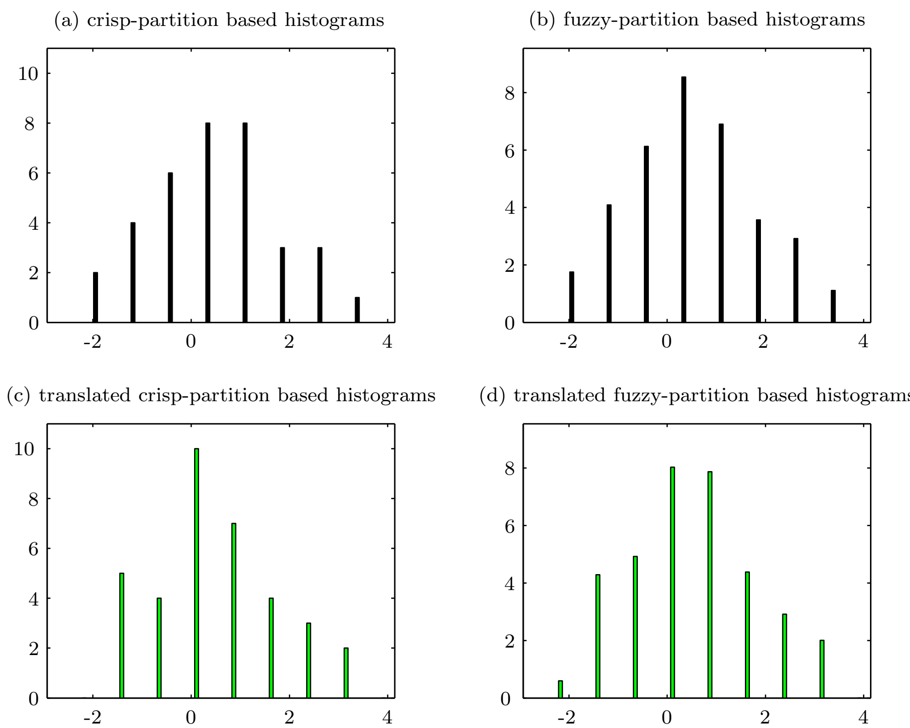

Fig. 2. Effect of the translation on a crisp ((a) and (b)) and a fuzzy ((c) and (d)) histogram

3 Fuzzy histogram density estimators

Expression (2) can be used for both crisp and fuzzy histograms to estimate the density underlying a set of observations. However, since A_k is a fuzzy subset, this expression no longer holds for any $x \in A_k$, but normalized accumulators $\frac{Acc_k}{nh}$ now have *degrees of truth* inherited from the fuzzy nature of A_k (see the preliminary of the section 1). The value $\frac{Acc_k}{nh}$ is then more *true* at m_k than at any other point of Ω. Our proposal is to assign this value $\frac{Acc_k}{nh}$ to the estimated density at each node m_k of the partition. Therefore, the estimated density can be obtained, at any point $x \neq m_k$, by interpolation.

In this paper, we propose to use, once again, the concept of strong uniform fuzzy partition of p fuzzy subsets to provide an interpolation of those p points.

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

$$\hat{f}_{FH}(x) = \frac{1}{nh} \sum_{k=1}^{p} Acc_k K_B(\frac{x - m_k}{h}) \tag{4}$$

where K_B is defined as in proposition 1 for the strong uniform fuzzy partition $(B_k)_{k=1,...,p}$.

Proof. Conditions 1 and 6 of the definition 1 imply that $\mu_{B_k}(m_l) = \delta_{kl}$ for $k, l \in \{1, ..., p\}$, where δ_{kl} is the Kronecker symbol, and, $K_B(\frac{m_l - m_k}{h}) = \mu_{B_k}(m_l)$. Then, $\hat{f}_{FH}(m_l) = \frac{Acc_l}{nh}$, for all $l \in \{1, ..., p\}$, which means that $\hat{f}_{FH}$ goes though 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 $(B_k)_{k=1,...,p}$, except at the nodes m_k, where the smoothness is not guaranteed. We can now add a convergence property of the estimators given by expression (4). So, let the error between the underlying density $f(x)$ and the estimate $\hat{f}_{FH}(x)$ be measured by the mean squared error : $MSE(x) \triangleq E_f[\hat{f}_{FH}(x) - f(x)]^2$. We have proved Theorem 1 in a paper to be published [LS], which is in some sense, the technical part of this paper. This proof is inspired from the demonstrations of the consistency theorems of the kernel density estimator, that are in [Tsy04].

Theorem 1. *Let us suppose*

1. *$f : \Omega \to [0, 1]$ is a density function such that f is bounded ($\forall x \in \Omega$, $f(x) \leq f_{max} < +\infty$) and f', its derivative, is bounded ($\forall x \in \Omega$, $|f'(x)| \leq f'_{max} < +\infty$),*
2. *K_A, as defined in proposition 1, verifies $\int_{-1}^{1} K_A^2(u)du < +\infty$.*

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

$$h \to 0 \text{ and } nh \to +\infty \;\Rightarrow\; MSE(x) \to 0 \tag{5}$$

This theorem gives a 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 ($h \to 0$ or $p \to +\infty$), and the growth of the mean number of data in each accumulator ($nh \to +\infty$).

However, the use of the membership functions of a fuzzy partition as interpolation functions is not compulsory. Thus, well-known interpolation functions could be used, e.g. the polynomial interpolation (with the Lagrange or the Newton form), or the spline interpolation, which improves the smoothness management at the nodes.

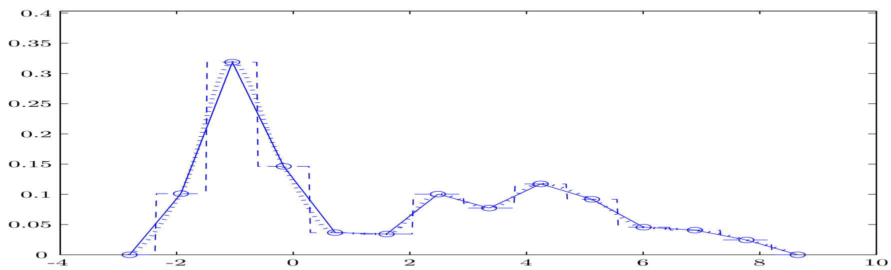

Fig. 3. Density estimation by interpolation

Figure 3 shows four estimations of a bimodal gaussian distribution with parameters $(m_1 = 4, \sigma_1^2 = 4)$ and $(m_2 = -1, \sigma_2^2 = \frac{1}{4})$, based upon a fuzzy triangular histogram. The circles are the interpolation points $(m_k, \frac{Acc_k}{nh})$. The dashed line is the crisp interpolation (see expression (2)). The solid line is the estimator obtained by fuzzy triangular interpolation (see expression (4)). The dotted line is a spline interpolation of the points $(m_k, \frac{Acc_k}{nh})$.

The estimations are obtained with $n = 100$ observations and $p = 14$ fuzzy subsets of the partition, which means that we are no longer in convergence conditions. Table 2 gives the empirical L_1 errors of interpolation, i.e. $\int_{\Omega} |\hat{f}_{FH}(x) - f(x)|dx$, obtained by repeating the experiment 100 times. This error is noted $m_{Error} \pm 3 * \sigma_{Error}$, where m_{Error} is the mean of the L_1 error over the 100 experiments and σ_{Error} its standard deviation.

Note that, whatever the interpolation scheme, compared to crisp histogram density estimators, the fuzzy histogram density estimators seem to be more stable (which can be measured by means of the standard deviation) and closer (in L_1 distance) to the underlying density (in that particular case).

Table 2. L_1 errors of interpolation

	crisp accumulators	fuzzy accumulators
crisp interpolation	0.028307 ± 0.014664	0.024957 ± 0.01166
triangular interpolation	0.021579 ± 0.014963	0.020807 ± 0.013539
cosine interpolation	0.022524 ± 0.014832	0.021103 ± 0.01327
spline interpolation	0.021813 ± 0.01577	0.020226 ± 0.0139
Lagrange interpolation	3.7349 ± 7.893	2.2537 ± 4.3761

Another important remark, deduced from Table 2, is that the fuzzy interpolants appear to be a good choice, because their error magnitudes are equivalent to those of the spline interpolant, which is known as being an optimal tool.

4 Conclusion

In this paper, we have presented density estimators based upon a fuzzy histogram. This latter being nothing else but a generalization of the popular crisp histogram, when replacing the crisp partition by a fuzzy partition. Those proposed density estimators consist in interpolations of the nodes' values of the density obtained in the usual way : $\frac{Acc_k}{nh}$.

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.

[Lee02] T. Leephakpreeda. Novel determination of differential-equation solutions : universal approximation method. *Journal of computational and applied mathematics*, 146:443–457, 2002.

[LS] K. Loquin and O. Strauss. Histogram density estimators based upon a fuzzy partition. *Laboratoire d´Informatique, de Robotique et de Microélectronique de Montpellier, Université Montpellier II.*

[Per04] I. Perfilieva. Fuzzy function as an approximate solution to a system of fuzzy relation equations. *Fuzzy sets and systems*, 147:363–383, 2004.

[Per06] I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy sets and systems*, 157:993–1023, 2006.

[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 quasicontinu. *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.

[Tsy04] A. B. Tsybakov. *Introduction à l'estimation non-paramétrique*. Springer-Verlag, 2004.

[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.