Fuzzy Histograms: A Statistical Analysis

Ludo Waltman, Uzay Kaymak, and Jan van den Berg

Faculty of Economics, Erasmus University Rotterdam

P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

lwaltman@ieee.org, u.kaymak@ieee.org, jvandenberg@few.eur.nl

Abstract

Fuzzy histograms are a fuzzy generalization of ordinary crisp histograms. In this paper, fuzzy histograms are analyzed statistically and are compared with other nonparametric density estimators. It turns out that fuzzy histograms can be used to combine a high level of statistical efficiency with a high level of computational efficiency.

Keywords: Fuzzy histogram, Histogram, Density estimation, Probability density function.

1   Introduction

Fuzzy histograms (FHs), introduced in [2, 4], are a fuzzy generalization of ordinary crisp histograms. FHs are very similar to the double-kernel estimators discussed in [5]. In this paper, FHs are analyzed statistically and are compared with other nonparametric density estimators. The analysis is based on the criterion of integrated mean squared error, which is typically used in the literature on nonparametric density estimation [3]. In the analysis, it is shown that a high level of statistical efficiency can be obtained by using FHs. This is an important advantage in comparison with crisp histograms. FHs generally also have a high level of computational efficiency. It should be noted that the analysis in this paper is restricted to univariate density estimation.

The paper is organized as follows. In Section 2, FHs are discussed and the concept of a uniform partitioning is introduced. A statistical analysis of FHs is presented in Section 3. Section 4 provides a discussion in which FHs are compared with other nonparametric density estimators.

2   Fuzzy Histograms

Let $x_1, \ldots, x_n$ denote a random sample of size n from a distribution with probability density function (pdf) $f(x)$. A FH estimates $f(x)$ as follows

$$\hat{f}(x) = \sum_i \frac{p_i \mu_i(x)}{\int \mu_i(x) \, dx}, \qquad (1)$$

where p_i is given by

$$p_i = \frac{1}{n} \sum_{j=1}^{n} \mu_i(x_j). \qquad (2)$$

Note that in this paper the symbol $\int$ should be read as $\int_{-\infty}^{\infty}$. The membership functions (mfs) μ_i in (1) and (2) must describe a fuzzy partition, which means that they must satisfy

$$\sum_i \mu_i(x) = 1 \quad \forall x \in \mathbb{R}. \qquad (3)$$

If for all i and all x $\mu_i(x)$ equals 0 or 1, then an ordinary crisp histogram is obtained. For a more detailed discussion of the idea of a FH we refer to [2, 4].

The analysis of FHs in this paper is restricted to the special case in which the mfs μ_i constitute a uniform partitioning. This means that there must be an infinite number of mfs and that these mfs must be given by

$$\mu_i(x) = \mu\left(\frac{x}{h} - i\right) \quad \forall i \in \mathbb{Z}, \qquad (4)$$

where $h > 0$ is a smoothing parameter and μ is a mf that must satisfy

$$\int x\mu(x)\,dx = 0 \qquad (5)$$

and

$$\sum_{i \in \mathbb{Z}} \mu(x+i) = 1 \quad \forall x \in \mathbb{R}. \qquad (6)$$

Equation (6) ensures that the condition in (3) is satisfied. Also, it can be derived from (6) that

$$\int \mu(x)\,dx = 1. \qquad (7)$$

From (4) and (7) it follows that $\int \mu_i(x)\,dx = h$ for all $i \in \mathbb{Z}$. By using this result and by substituting (2) and (4) in (1) we obtain

$$\hat{f}(x) = \frac{1}{n}\sum_{j=1}^{n} K(x, x-x_j), \qquad (8)$$

where the function $K(x,w)$ is given by

$$K(x,w) = \frac{1}{h}\sum_{i \in \mathbb{Z}} \mu\left(\frac{x}{h}-i\right)\mu\left(\frac{x-w}{h}-i\right). \qquad (9)$$

Using (8) and (9), a FH can be seen as a generalization of a kernel density estimator (KDE) or, more specifically, as a KDE that uses different kernels for different values of x. Similarly to a KDE, a FH has one parameter, which is the smoothing parameter h. The optimal value of this parameter depends on the sample size n. In the next section, FHs are analyzed by extending the analysis of KDEs in [3].

3   Statistical Analysis

For analyzing nonparametric density estimators, the integrated mean squared error (IMSE, also called the mean integrated squared error or MISE) is usually considered [3]. The IMSE is defined as

$$\text{IMSE} = \int \text{E}\left(\hat{f}(x) - f(x)\right)^2 dx. \qquad (10)$$

The IMSE equals the sum of the integrated squared bias (ISB) and the integrated variance (IV), which are defined as

$$\text{ISB} = \int \left(\text{E}\hat{f}(x) - f(x)\right)^2 dx \qquad (11)$$

and

$$\text{IV} = \int \text{E}\left(\hat{f}(x) - \text{E}\hat{f}(x)\right)^2 dx. \qquad (12)$$

Usually, the asymptotically optimal IMSE, denoted by IMSE* in this paper, is derived. This is the IMSE that is obtained by choosing the parameters of a density estimator in an asymptotically optimal way (i.e. in such a way that the terms in the IMSE that are relevant in an asymptotic analysis are minimized).

The IMSE* of a FH is given in the following theorem.

Theorem 1 *Consider a FH. Let the mfs μ_i constitute a uniform partitioning. The IMSE* of the FH is then given by*

$$\text{IMSE}^* = 3\left(\frac{\alpha}{4}R\left(f'\right)\right)^{1/3}\gamma^{2/3}n^{-2/3} + O\left(n^{-1}\right) \qquad (13)$$

if $\alpha > 0$ and

$$\text{IMSE}^* = \frac{5}{4}\left(\beta R\left(f''\right)\right)^{1/5}\gamma^{4/5}n^{-4/5} + O\left(n^{-1}\right) \qquad (14)$$

if $\alpha = 0$, where

$$\alpha = \int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\sum_{i \in \mathbb{Z}}(x-i)\mu(x-i)\right)^2 dx, \qquad (15)$$

$$\beta = \int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\int w^2\mu(w)\,dw \right.$$
$$\left. + \sum_{i \in \mathbb{Z}}(x-i)^2\mu(x-i)\right)^2 dx, \qquad (16)$$

and

$$\gamma = \int_{-\frac{1}{2}}^{\frac{1}{2}}\int\left(\sum_{i \in \mathbb{Z}}\mu(x-i)\mu(w-i)\right)^2 dw\,dx. \qquad (17)$$

In (13) and (14), $R(\cdot)$ denotes the roughness of a function and is defined as [3]

$$R(\phi) = \int \phi(x)^2\,dx. \qquad (18)$$

The proof of Theorem 1 is given in the following three paragraphs. First, the ISB and the IV are derived in Paragraph 3.1 and 3.2, respectively. Then, the IMSE* is derived in Paragraph 3.3.

3.1 Integrated Squared Bias

Let X denote a random variable that is distributed according to $f(x)$. The expectation of a FH can then be written as

$$
\begin{aligned}
\mathrm{E}\hat{f}(x) &= \mathrm{E}K(x, x - X) \\
&= \int K(x, x - t)f(t)\, dt \\
&= \int K(x, w)f(x - w)\, dw \\
&= \int K(x, w)\sum_{k=0}^{\infty} \frac{(-w)^k f^{(k)}(x)}{k!}\, dw \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k f^{(k)}(x)}{k!} \int w^k K(x, w)\, dw. \quad (19)
\end{aligned}
$$

Note that this result is based on a Taylor series expansion. It further follows from (6), (7), and (9) that $\int K(x, w)\, dw = 1$ for all $x \in \mathbb{R}$. The bias $B(x)$ of a FH therefore equals

$$
\begin{aligned}
B(x) &= \mathrm{E}\hat{f}(x) - f(x) \\
&= \sum_{k=1}^{\infty} \frac{(-1)^k f^{(k)}(x)}{k!} \int w^k K(x, w)\, dw. \quad (20)
\end{aligned}
$$

Using (9), the integral in (19) and (20) can be written as

$$
\begin{aligned}
&\int \frac{w^k}{h} \sum_{i\in\mathbb{Z}} \mu\left(\frac{x}{h} - i\right) \mu\left(\frac{x - w}{h} - i\right) dw \\
&= \sum_{i\in\mathbb{Z}} \mu\left(\frac{x}{h} - i\right) \int \frac{w^k}{h} \mu\left(\frac{x - w}{h} - i\right) dw \\
&= \sum_{i\in\mathbb{Z}} \mu\left(\frac{x}{h} - i\right) \int (x - hi - hv)^k \mu(v)\, dv.
\end{aligned}
$$
$$(21)$$

The ISB of a FH follows from (20) and is given by

$$
\begin{aligned}
\mathrm{ISB} &= \int B(x)^2\, dx \\
&= h \int \left(\sum_{k=1}^{\infty} \frac{(-1)^k f^{(k)}(hu)}{k!} \right. \\
&\quad \left. \times \int w^k K(hu, w)\, dw \right)^2 du. \quad (22)
\end{aligned}
$$

Substitution of hu for x in (21) results in

$$
h^k \sum_{i\in\mathbb{Z}} \mu(u - i) \int (u - i - v)^k \mu(v)\, dv. \quad (23)
$$

Using (5) and (7), for $k = 1$ (23) becomes

$$
h \sum_{i\in\mathbb{Z}} (u - i)\mu(u - i). \quad (24)
$$

Assuming that the expression in (24) does not equal 0, it follows from (22), (23), and (24) that the ISB of a FH can be written as

$$
\begin{aligned}
&h \int \left(-hf'(hu) \sum_{i\in\mathbb{Z}} (u - i)\mu(u - i) + \ldots \right)^2 du \\
&= h^3 \int f'(hu)^2 \left(\sum_{i\in\mathbb{Z}} (u - i)\mu(u - i) \right)^2 du + \ldots \\
&= h^3 \sum_{j\in\mathbb{Z}} \int_{j-\frac{1}{2}}^{j+\frac{1}{2}} f'(hu)^2 \\
&\quad \times \left(\sum_{i\in\mathbb{Z}} (u - i)\mu(u - i) \right)^2 du + \ldots \\
&= \alpha h^3 \sum_{j\in\mathbb{Z}} f'(h\eta_j)^2 + \ldots, \quad (25)
\end{aligned}
$$

where the last step uses the generalized mean value theorem. This step is valid for some collection of points η_j, where $j - \frac{1}{2} < \eta_j < j + \frac{1}{2}$. The generalized mean value theorem states that

$$
\int_a^b \phi(x)g(x)\, dx = \phi(c) \int_a^b g(x)\, dx, \quad (26)
$$

for some value of c such that $a < c < b$. The functions ϕ and g are assumed to be continuous on the interval $[a, b]$, and g is also assumed to be nonnegative on this interval. The last step in (25) also uses the following result

$$
\int_{a-\frac{1}{2}}^{a+\frac{1}{2}} \left(\sum_{i\in\mathbb{Z}} (x - i)\mu(x - i) \right)^2 dx = \alpha \quad \forall a \in \mathbb{R}. \quad (27)
$$

Due to space limitations we do not prove this result. Using numerical integration approximations, (25) can be written as

$$
\mathrm{ISB} = \alpha R\left(f'\right) h^2 + O\left(h^3\right). \quad (28)
$$

It can be seen that in the special case where $\alpha = 0$ the expression in (24) also equals 0. We derive a separate expression for the ISB of a FH in this special case. Using (5), (6), and (7), for $k = 2$ (23) becomes

$$h^2 \left(\int v^2 \mu(v) \, dv + \sum_{i \in \mathbb{Z}} (u-i)^2 \mu(u-i) \right). \quad (29)$$

In a similar way as above, it then follows that the ISB of a FH in (22) can be written as

$$h \int \left(\left(\frac{1}{2} h^2 f''(hu) \left(\int v^2 \mu(v) \, dv \right. \right. \right.$$

$$\left. \left. \left. + \sum_{i \in \mathbb{Z}} (u-i)^2 \mu(u-i) \right) + \dots \right)^2 du \right.$$

$$= \frac{\beta}{4} h^5 \sum_{j \in \mathbb{Z}} f''(h\eta_j)^2 + \dots, \quad (30)$$

where the last step, which is valid for some collection of points η_j $(j - \frac{1}{2} < \eta_j < j + \frac{1}{2})$, uses the generalized mean value theorem. Rewriting (30) using numerical integration approximations results in the following expression for the ISB of a FH in the special case where $\alpha = 0$

$$\text{ISB} = \frac{\beta}{4} R\left(f''\right) h^4 + O\left(h^5\right). \quad (31)$$

For convergence (in mean square) of the FH estimator $\hat{f}(x)$ to the pdf $f(x)$ that is being estimated, it is necessary that ISB $\to 0$ as $n \to \infty$. It follows from both (28) and (31) that, in order to achieve this, the smoothing parameter h must be chosen in such a way that $h \to 0$ as $n \to \infty$.

3.2 Integrated Variance

The variance $V(x)$ of a FH equals

$$V(x) = \text{E}\left(\hat{f}(x) - \text{E}\hat{f}(x) \right)^2$$

$$= \frac{1}{n} \text{E}\left(K(x, x-X)^2 \right) - \frac{1}{n} (\text{E}K(x, x-X))^2. \quad (32)$$

The first term in (32) can be written as

$$\frac{1}{n} \int K(x, x-t)^2 f(t) \, dt$$

$$= \frac{1}{n} \int K(x, w)^2 f(x-w) \, dw$$

$$= \frac{1}{n} \int K(x, w)^2 \sum_{k=0}^{\infty} \frac{(-w)^k f^{(k)}(x)}{k!} \, dw$$

$$= \frac{1}{n} \sum_{k=0}^{\infty} \frac{(-1)^k f^{(k)}(x)}{k!} \int w^k K(x, w)^2 \, dw. \quad (33)$$

Using (9), the integral in (33) becomes

$$\int \frac{w^k}{h^2} \left(\sum_{i \in \mathbb{Z}} \mu\left(\frac{x}{h} - i\right) \mu\left(\frac{x-w}{h} - i\right) \right)^2 dw$$

$$= \frac{1}{h} \int (x-hv)^k \left(\sum_{i \in \mathbb{Z}} \mu\left(\frac{x}{h} - i\right) \mu(v-i) \right)^2 dv. \quad (34)$$

The IV of a FH follows from (32) and is given by

$$\text{IV} = \int V(x) \, dx = h \int V(hu) \, du$$

$$= \frac{h}{n} \int \text{E}\left(K(hu, hu - X)^2 \right) du$$

$$- \frac{h}{n} \int (\text{E}K(hu, hu - X))^2 \, du. \quad (35)$$

Substitution of hu for x in (34) results in

$$h^{k-1} \int (u-v)^k \left(\sum_{i \in \mathbb{Z}} \mu(u-i)\mu(v-i) \right)^2 dv. \quad (36)$$

Using (33) and (36), the first term in (35) can be written as

$$\frac{1}{n} \sum_{k=0}^{\infty} \frac{(-h)^k}{k!} \int f^{(k)}(hu) \int (u-v)^k$$

$$\times \left(\sum_{i \in \mathbb{Z}} \mu(u-i)\mu(v-i) \right)^2 dv \, du$$

$$= \frac{1}{n} \int f(hu)$$

$$\times \int \left(\sum_{i \in \mathbb{Z}} \mu(u-i)\mu(v-i) \right)^2 dv \, du + \dots$$

$$= \frac{\gamma}{n} \sum_{j \in \mathbb{Z}} f(h\eta_j) + \dots, \quad (37)$$

where the last step, which is valid for some collection of points η_j $(j - \frac{1}{2} < \eta_j < j + \frac{1}{2})$, uses the generalized mean value theorem. Rewriting (37) using numerical integration approximations gives

$$\frac{\gamma}{hn} + O\left(n^{-1}\right). \tag{38}$$

To see this, note that $\int f(x)\,dx = 1$ since $f(x)$ is a pdf. Using (6), (7), (19), and (23), the second term in (35) can be written as

$$\frac{h}{n}\int\left(\sum_{k=0}^{\infty}\frac{(-1)^k f^{(k)}(hu)}{k!}\right.$$
$$\left.\times\int w^k K(hu,w)\,dw\right)^2 du$$
$$= \frac{h}{n}\int f(hu)^2\,du + \ldots. \tag{39}$$

In a similar way as the first term in (35), the second term can be rewritten using the generalized mean value theorem and numerical integration approximations. Equation (39) then becomes

$$\frac{R(f)}{n} + O\left(hn^{-1}\right). \tag{40}$$

The IV equals the sum of (38) and (40). It was discussed above that for convergence of $\hat{f}(x)$ to $f(x)$ the smoothing parameter h must be chosen in such a way that $h \to 0$ as $n \to \infty$. Using such a choice of h, (38) decreases at a lower rate than (40) as $n \to \infty$. The IV can therefore simply be written as

$$\text{IV} = \frac{\gamma}{hn} + O\left(n^{-1}\right). \tag{41}$$

3.3 Asymptotically Optimal Integrated Mean Squared Error

The IMSE equals the sum of the ISB and the IV. For $\alpha > 0$, this means that the IMSE equals the sum of (28) and (41), which results in

$$\text{IMSE} = \alpha R\left(f'\right)h^2 + \frac{\gamma}{hn} + O\left(h^3 + n^{-1}\right). \tag{42}$$

The asymptotically optimal smoothing parameter h^* is the smoothing parameter h that minimizes the first two terms in (42). This gives

$$h^* = \left(\frac{\gamma}{2\alpha R\left(f'\right)}\right)^{1/3}n^{-1/3}. \tag{43}$$

By substituting (43) in (42) the IMSE* in (13) is obtained.

For $\alpha = 0$, the IMSE equals the sum of (31) and (41). This results in

$$\text{IMSE} = \frac{\beta}{4}R\left(f''\right)h^4 + \frac{\gamma}{hn} + O\left(h^5 + n^{-1}\right). \tag{44}$$

The asymptotically optimal smoothing parameter h^* is then given by

$$h^* = \left(\frac{\gamma}{\beta R\left(f''\right)}\right)^{1/5}n^{-1/5}. \tag{45}$$

By substituting (45) in (44) the IMSE* in (14) is obtained.

Theorem 1 has now been proven. We briefly discuss the choice of mf μ in the next paragraph.

3.4 Choice of the Membership Function

It follows from Theorem 1 that the choice of mf μ strongly affects the statistical efficiency of a FH. If μ is chosen in such a way that $\alpha > 0$, then the IMSE* of a FH decreases at a rate of $O\left(n^{-2/3}\right)$. If, on the other hand, μ is chosen in such a way that $\alpha = 0$, then a much better rate of $O\left(n^{-4/5}\right)$ is obtained. To maximize the efficiency of a FH, it is therefore essential to know what choices of μ result in $\alpha = 0$.

It can be shown that there are several choices of μ for which $\alpha = 0$. As an example of such a choice, we consider the following triangular mf

$$\mu(x) = \max(1 - |x|, 0). \tag{46}$$

For μ given by (46), it can be calculated that $\beta = 7/60$ and $\gamma = 1/2$. Using (14), we then obtain

$$\text{IMSE}^* = \frac{5}{8}\left(\frac{7}{30}R\left(f''\right)\right)^{1/5}n^{-4/5} + O\left(n^{-1}\right). \tag{47}$$

We note that in [1] a density estimator is studied that is equivalent to a FH with μ given by (46).

4 Discussion

Two important criteria for choosing a nonparametric density estimator are statistical efficiency and computational efficiency. Using these criteria, ordinary crisp histograms and KDEs are

two extremes. Crisp histograms are very efficient computationally but quite inefficient statistically, whereas KDEs are quite efficient statistically but very inefficient computationally [3]. The difference in statistical efficiency follows from the IMSE* criterion. The IMSE* of a crisp histogram decreases at a rate of $O\left(n^{-2/3}\right)$, whereas the IMSE* of a KDE decreases at a rate of $O\left(n^{-4/5}\right)$.

According to Theorem 1, the rate at which the IMSE* of a FH decreases depends on the value of α. For $\alpha > 0$, the IMSE* decreases at a rate of $O\left(n^{-2/3}\right)$, which is quite inefficient. For $\alpha = 0$, on the other hand, the IMSE* decreases at a quite efficient rate of $O\left(n^{-4/5}\right)$. Therefore, by choosing an appropriate mf μ, like the triangular mf in (46), a FH can be a statistically quite efficient density estimator. Since in general a FH can also be shown to be quite efficient computationally (we do not elaborate the issue of computational efficiency in this paper), both a high level of statistical efficiency and a high level of computational efficiency can be obtained by using a FH. This means that a FH makes it possible to combine the advantages of crisp histograms and KDEs into a single density estimator.

In the literature on nonparametric density estimation [3], other density estimators are described that also combine a high level of statistical efficiency with a high level of computational efficiency. An example is the frequency polygon (FP), which is obtained by linearly interpolating adjacent mid-bin values of a crisp histogram. Another example is the combination of a FP and an averaged shifted histogram (ASH). An ASH results from averaging a number of crisp histograms with the same bin width but different locations of the bin edges. Like KDEs and some FHs, FPs and FP-ASH combinations have an IMSE* that decreases at a rate of $O\left(n^{-4/5}\right)$ [3].

In Table 1, we compare the statistical efficiency of some nonparametric density estimators. All estimators in the table have an IMSE* that decreases at a rate of $O\left(n^{-4/5}\right)$. The KDE uses the optimal Epanechnikov kernel [3], while the FH uses a triangular mf μ given by (46). The ASH in the FP-ASH combination results from averaging two crisp histograms. An estimator's equivalent sam-

Table 1: Comparison of density estimators.

Density estimator	Equivalent sample size
KDE	1.000
FH	1.089
FP	1.269
FP-ASH	1.078

ple size is defined as the ratio of the sample size required by the estimator and the sample size required by a KDE with the Epanechnikov kernel such that they both have the same main term in their IMSE*. The equivalent sample sizes in Table 1 were calculated using results from [3]. The table shows that asymptotically the FH requires only 8.9% more data than the KDE to obtain the same IMSE*. The FH turns out to be considerably more efficient than the FP and slightly less efficient than the FP-ASH combination.

It is important to note that from the point of view of statistical efficiency the triangular mf in (46) need not be the optimal choice of the mf μ that is used in a FH. Finding the optimal choice of μ remains for future research.

References

[1] M.C. Jones. Discretized and interpolated kernel density estimates. *J. Am. Stat. Assoc.*, 84:733–741, 1989.

[2] U. Kaymak, W.-M. van den Bergh, and J. van den Berg. A fuzzy additive reasoning scheme for probabilistic Mamdani fuzzy systems. In *Proc. 12th IEEE Int. Conf. Fuzzy Syst.*, pages 331–336, 2003.

[3] D.W. Scott. *Multivariate density estimation.* John Wiley & Sons, 1992.

[4] J. van den Berg, U. Kaymak, and W.-M. van den Bergh. Financial markets analysis by using a probabilistic fuzzy modelling approach. *Int. J. Approx. Reason.*, 35:291–305, 2004.

[5] G.C. van den Eijkel. *Fuzzy probabilistic learning and reasoning.* PhD thesis, Delft University of Technology, 1998.