Metrika (2012) 75:5–22
DOI 10.1007/s00184-010-0311-y

# Fuzzy density estimation

**Mohsen Arefi · Reinhard Viertl · S. Mahmoud Taheri**

Received: 1 July 2009 / Published online: 29 April 2010


**Abstract** A new approach to density estimation with fuzzy random variables (FRV) is developed. In this approach, three methods (histogram, empirical c.d.f., and kernel methods) are extended for density estimation based on $\alpha$-cuts of FRVs.

**Keywords** Fuzzy density estimation · Fuzzy random variable (FRV) · Empirical cumulative distribution function (c.d.f.) · Histogram method · Kernel method

## 1 Introduction and preliminaries

In many applications it is necessary to estimate a probability density function (p.d.f.), based on a set of data. These data may be precise or non-precise quantities. Numerous methods for constructing p.d.f. have been proposed by different authors, using precise data.

Historically, the topic of density estimation has been investigated over the last few decades. Rosenblatt (1956), Rosenblatt (1971), Parzen (1962), Prasaka Rao (1983), Devroye and Györfi (1985), Silverman (1986), Jones (1991) and Lee et al. (2004) studied the problem of probability density estimation from different perspectives. Hazelton

M. Arefi (✉) · S. M. Taheri
Department of Mathematical Sciences, Isfahan University of Technology, Isfahan 84156-83111, Iran
e-mail: Arefi@math.iut.ac.ir

S. M. Taheri
e-mail: Taheri@cc.iut.ac.ir

R. Viertl
Technische Universität Wien, Institut für Statistik und Wahrscheinlichkeitstheorie,
Wiedner Hauptstrasse 8-10/107, 1040 Wien, Austria
e-mail: R.Viertl@tuwien.ac.at

(2000) investigated a marginal density estimation based on incomplete bivariate data. Fokianos (2004) and Qin and Zhang (2005) have adopted a quite different approach, using an $m$-sample density ratio model. They first estimated the parameters by maximizing a semiparametric likelihood function, and then they obtained the maximum semiparametric likelihood estimator of the unknown distribution function by putting weights on all the observations. Inference for parameters of the density ratio model in the case $m = 2$ has been studied by Qin (1998), Cheng and Chu (2004) and Keziou and Leoni-Aubin (2005, 2007). Also, Alberts and Karunamuni (2003), Qin and Zhang (2005) and Wu et al. (2007) used an iterative approach to select the bandwidth and established asymptotic normality of certain estimators. Cheng (1994) and Cheng and Chu (1996) investigated the problem of density estimation with missing data. For a comprehensive review on smoothing methods, see Wand and Jones (1995) and Simonoff (1996). Viertl (1996) and Loquin and Strauss (2008) studied some methods for density estimation with non-precise (fuzzy) information. Also, Trutschnig (2008), using an analogous to Dempster theory of interval-valued probabilities, investigated the strong consistency of fuzzy relative frequencies interpreted as estimator for the fuzzy-valued probability. For more studies about fuzzy statistics and fuzzy data see Kruse and Meyer (1987), Viertl (1996), Viertl (2006) and Taheri (2003).

This paper provides a description of probability density estimation (p.d.f) with fuzzy random variables. We extend three methods (histogram, empirical c.d.f., and kernel) for density estimation using $\alpha$-cuts of fuzzy data.

The organization of this paper is as follows. In Sect. 2, we review some preliminary concepts for density estimation. In Sect. 3, the histogram method is generalized for estimating density function to the case when the statistical data are fuzzy rather than crisp. A procedure based on empirical c.d.f. method for estimating fuzzy density function is provided in Sect. 4. Section 5 includes fuzzy density estimation based on kernel method. In Sect. 6, we investigate the three proposed methods by a practical example. Finally, Sect. 7 provides an overall conclusion.

Let us first recall some preliminary concepts. A fuzzy set $\widetilde{A}$ of a universe $\Omega$ is defined by a membership function $\widetilde{A} : \Omega \to [0, 1]$. An $\alpha$-cut of $\widetilde{A}$, denoted by $\widetilde{A}[\alpha]$, is defined as $\widetilde{A}[\alpha] = \{x|\widetilde{A}(x) \geq \alpha\}$, for $0 < \alpha \leq 1$.
A fuzzy number $\widetilde{N}$ is a fuzzy set of the real numbers satisfying:

(i) $\exists\, x \in \Re$ such that $\widetilde{N}[1] = \{x\}$,
(ii) $\widetilde{N}[\alpha]$ is a closed bounded interval for all $0 < \alpha \leq 1$.

The $\alpha$-cut of each fuzzy number $\widetilde{N}$ is usually denoted by $\widetilde{N}[\alpha] = \left[\widetilde{N}^L[\alpha], \widetilde{N}^U[\alpha]\right]$, where $\widetilde{N}^L[\alpha] = \inf\{x|\widetilde{N}(x) \geq \alpha\}$ and $\widetilde{N}^U[\alpha] = \sup\{x|\widetilde{N}(x) \geq \alpha\}$ (see Klir and Yuan 1995).
A triangular fuzzy number $\widetilde{T} = (a, s_1, s_2)_T$ is defined as

$$\tilde{T}(x) = \begin{cases} 1 + \frac{x-a}{s_1} & a - s_1 \leq x < a, \\ 1 + \frac{a-x}{s_2} & a \leq x < a + s_2, \\ 0 & \text{otherwise.} \end{cases}$$

Let $I = [a, b]$ and $J = [c, d]$ be two closed intervals. Then, based on the interval arithmetic (Klir and Yuan 1995), we have

$$I + J = [a + c, b + d],$$
$$I - J = [a - d, b - c],$$
$$I.J = [\alpha_1, \beta_1], \quad \alpha_1 = \min\{ac, ad, bc, bd\}, \quad \beta_1 = \max\{ac, ad, bc, bd\},$$
$$I \div J = [\alpha_2, \beta_2], \quad \alpha_2 = \min\left\{\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right\}, \quad \beta_2 = \max\left\{\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right\},$$

where zero does not belong to $J = [c, d]$ in the last case.

Let $\mathcal{F}(\Re)$ be the set of all fuzzy numbers. Given a probability space $(\Omega, \mathcal{A}, P)$, a *fuzzy random variable* (FRV) is a Borel measurable mapping $\widetilde{\mathcal{X}} : \Omega \to \mathcal{F}(\Re)$ such that for any $\alpha \in [0, 1]$, the $\alpha$-cut $\widetilde{\mathcal{X}}[\alpha]$ is a random interval, i.e. $\widetilde{\mathcal{X}}[\alpha] : \Omega \to \mathcal{H}(\Re)$ ($\mathcal{H}(\Re)$ is the class of nonempty compact intervals) (see Puri and Ralescu 1986; Gil 2004). It is Borel measurable with respect to the Borel $\sigma$-field generated by the topology associated with the Hausdorff metric on $\mathcal{H}(\Re)$,

$$d_H(A, B) = \max\left\{\sup_{a\in A}\inf_{b\in B}|a - b|, \sup_{b\in B}\inf_{a\in A}|a - b|\right\}.$$

Let $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$ be two fuzzy random variables. We say that $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$ are independent if and only if each random variable in the set $\left\{\widetilde{\mathcal{X}}^L[\alpha], \widetilde{\mathcal{X}}^U[\alpha] : \alpha \in [0, 1]\right\}$ is independent of each random variable in the set $\left\{\widetilde{\mathcal{Y}}^L[\alpha], \widetilde{\mathcal{Y}}^U[\alpha] : \alpha \in [0, 1]\right\}$. $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$ are called identically distributed if $\widetilde{\mathcal{X}}^L[\alpha]$ and $\widetilde{\mathcal{Y}}^L[\alpha]$, and also $\widetilde{\mathcal{X}}^U[\alpha]$ and $\widetilde{\mathcal{Y}}^U[\alpha]$ are identically distributed for all $\alpha \in [0, 1]$. Similar arguments can be used for more than two FRVs.

Hence, we say that the fuzzy random variables $\widetilde{\mathcal{X}}_1, \ldots, \widetilde{\mathcal{X}}_n$ are i.i.d. (written as i.i.d. FRV) if they are independent and identically distributed (for more details, see Klement et al. 1986; Kruse and Meyer 1987; Li and Ogura 2006).

## 2 Density estimation: an overall review

For our purposes in this article, let us briefly overview density estimation methods (for more details, see, for example Prasaka Rao 1983; Owen 2001; Shao 2003). Let $X = (X_1, \ldots, X_n)$ be a random sample with observed values $x = (x_1, \ldots, x_n)$, from a population with probability density function (or probability mass function) $f_X(.)$. Density estimation is the attempt to either parametrically or nonparametrically approximate the p.d.f. of $X$. In parametric density estimation, the focus is on fitting a probability distribution to the collected data. This approach requires to determine the appropriate theoretical distribution and its corresponding parameters. Nonparametric density estimation seeks to allow the data to stand on their own by using alternative methods to determine the form of the density function. Three major nonparametric approaches for density estimation are the histogram method, the empirical cumulative distribution function (c.d.f.) method, and the kernel method.

All these methods have a similar basic idea: The p.d.f. of the random variable $X$ may be defined as

$$f_X(t) = \frac{d}{dx}F_X(t) = \lim_{h_n\to 0}\frac{F(t + h_n) - F(t - h_n)}{2h_n}. \tag{1}$$

Now, given a random sample $X_1, \ldots, X_n$ from $f_X(.)$, an approximation for the density function (1) may be obtained by allowing $h_n$ to be very small and using the relative frequency function for the numerator, yielding

$$\hat{f}_n(t) = \frac{F_n(t+h_n) - F_n(t-h_n)}{2h_n} = \frac{1}{2nh_n} \sum_{i=1}^{n} I_{(t-h_n,\ t+h_n]}(x_i), \tag{2}$$

where, $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,t]}(x_i)$ is the empirical c.d.f., and $\{h_n\}$ is a decreasing sequence of positive constants.

## 3 Fuzzy density estimation based on histogram method

Suppose that we have i.i.d. FRVs $\widetilde{\mathcal{X}}_1, \ldots, \widetilde{\mathcal{X}}_n$ from a population with p.d.f. $f_X(.)$, and we observe the fuzzy values $\widetilde{X}_1, \ldots, \widetilde{X}_n$ (see Puri and Ralescu 1986). In this section, we define a method to obtain a fuzzy density estimation, based on an extended histogram method.

**Definition 1** Suppose that a fuzzy sample $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$ from a population with unknown p.d.f. $f_X(\cdot)$, whose domain is $[a, b]$, is observed. Let $\widetilde{X}_i[\alpha]$ be the $\alpha$-cut of $\widetilde{X}_i$ obeying $\widetilde{X}_i[\alpha] = [\widetilde{X}_i^L[\alpha], \widetilde{X}_i^U[\alpha]]$, $i = 1, \ldots, n$. Then for partition $a = t_0, t_1, \ldots, t_{m-1}, t_m = b$ the fuzzy histogram is defined by the fuzzy values $\hat{f}_H(t)$ whose $\alpha$-cuts $\hat{f}_H(t)[\alpha]$ are given by

$$\begin{cases} \hat{f}_H(t)[\alpha] = c_j[\alpha], & \text{for } t_j \le t < t_{j+1}, \quad j = 0, 1, \ldots, m-1, \\ \hat{f}_H(b)[\alpha] = c_{m-1}[\alpha], & \\ \hat{f}_H(t)[\alpha] = 0, & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} c_j[\alpha] &= \frac{1}{n(t_{j+1}-t_j)} \sum_{i=1}^{n} \left[\inf\left(I_{[t_j,t_{j+1})}(y);\ y \in \widetilde{X}_i[\alpha]\right), \sup\left(I_{[t_j,t_{j+1})}(y);\ y \in \widetilde{X}_i[\alpha]\right)\right] \\ &= \left[\frac{1}{n(t_{j+1}-t_j)} \sum_{i=1}^{n} \inf\left(I_{[t_j,t_{j+1})}(y);\ y \in \widetilde{X}_i[\alpha]\right),\right. \\ &\qquad \left.\frac{1}{n(t_{j+1}-t_j)} \sum_{i=1}^{n} \sup\left(I_{[t_j,t_{j+1})}(y);\ y \in \widetilde{X}_i[\alpha]\right)\right], \end{aligned}$$

in which

$$I_{[t_j,t_{j+1})}(y) = \left\{\begin{matrix} 1 \text{ for } t_j \le y < t_{j+1} \\ 0 \text{ otherwise} \end{matrix}\right\} \quad \forall\, y \in \Re, \quad j = 0, 1, \ldots, m-1.$$

*Example 1* Suppose that, based on a fuzzy random sample of size $n = 30$, we obtain the following triangular fuzzy numbers (data are taken from Viertl and Hareter 2006):

| No. | Data | No. | Data | No. | Data |
|---|---|---|---|---|---|
| 1 | $(0.23, 0.04, 0.07)_T$ | 11 | $(0.41, 0.03, 0.08)_T$ | 21 | $(0.64, 0.11, 0.07)_T$ |
| 2 | $(0.76, 0.05, 0.02)_T$ | 12 | $(0.86, 0.08, 0.04)_T$ | 22 | $(0.94, 0.09, 0.04)_T$ |
| 3 | $(0.98, 0.12, 0.09)_T$ | 13 | $(1.02, 0.03, 0.10)_T$ | 23 | $(1.08, 0.10, 0.06)_T$ |
| 4 | $(1.14, 0.06, 0.09)_T$ | 14 | $(1.23, 0.03, 0.14)_T$ | 24 | $(1.37, 0.08, 0.06)_T$ |
| 5 | $(1.46, 0.10, 0.07)_T$ | 15 | $(1.53, 0.13, 0.15)_T$ | 25 | $(1.64, 0.02, 0.08)_T$ |
| 6 | $(1.69, 0.05, 0.12)_T$ | 16 | $(1.78, 0.04, 0.06)_T$ | 26 | $(1.83, 0.09, 0.05)_T$ |
| 7 | $(1.95, 0.05, 0.11)_T$ | 17 | $(1.99, 0.08, 0.09)_T$ | 27 | $(2.04, 0.11, 0.06)_T$ |
| 8 | $(2.17, 0.03, 0.05)_T$ | 18 | $(2.25, 0.04, 0.04)_T$ | 28 | $(2.36, 0.05, 0.09)_T$ |
| 9 | $(2.40, 0.08, 0.12)_T$ | 19 | $(2.45, 0.01, 0.08)_T$ | 29 | $(2.49, 0.13, 0.05)_T$ |
| 10 | $(2.51, 0.10, 0.14)_T$ | 20 | $(2.57, 0.07, 0.02)_T$ | 30 | $(2.61, 0.08, 0.06)_T$ |

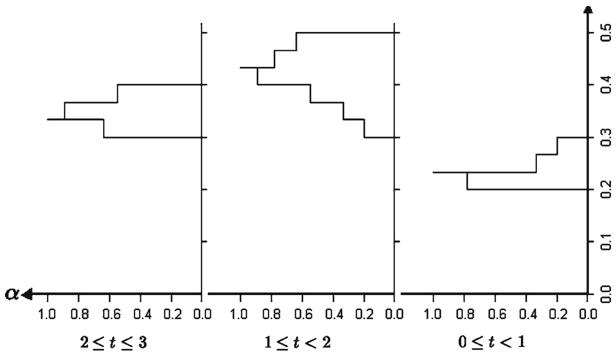


**Fig. 1** Fuzzy density estimation based on fuzzy histogram method, in Example 1

**Table 1** Some $\alpha$-cuts of the fuzzy density estimation, based on histogram method, in Example 1

| | $0 \le t < 1$ | $1 \le t < 2$ | $2 \le t \le 3$ |
|---|---|---|---|
| $\alpha = 0.0$ | [0.2000, 0.3000] | [0.3000, 0.5000] | [0.3000, 0.4000] |
| $\alpha = 0.2$ | [0.2000, 0.2667] | [0.3333, 0.5000] | [0.3000, 0.4000] |
| $\alpha = 0.4$ | [0.2000, 0.2333] | [0.3667, 0.5000] | [0.3000, 0.4000] |
| $\alpha = 0.6$ | [0.2000, 0.2333] | [0.4000, 0.5000] | [0.3000, 0.3667] |
| $\alpha = 0.8$ | [0.2333, 0.2333] | [0.4000, 0.4333] | [0.3333, 0.3667] |
| $\alpha = 1.0$ | [0.2333, 0.2333] | [0.4333, 0.4333] | [0.3333, 0.3333] |

The fuzzy density estimation based on histogram method is shown in Fig. 1. Hence, the fuzzy density estimation is approximately $\frac{7}{30}$ for every $t \in [0, 1)$, approximately $\frac{13}{30}$ for every $t \in [1, 2)$, and approximately $\frac{1}{3}$ for every $t \in [2, 3]$. Also, some $\alpha$-cuts of the related results are shown in Table 1.

For instance, the estimate of the density for the value t=2.5, $\hat{f}_H(t = 2.5)$ is "approximately $\frac{1}{3}$", which is described qualitatively by a fuzzy number. Some $\alpha$-cuts of this fuzzy number are given in Table 1.

## 4 Fuzzy density estimation based on empirical c.d.f. method

Suppose that we have i.i.d. FRVs $\widetilde{\mathcal{X}}_1, \ldots, \widetilde{\mathcal{X}}_n$ from a population with p.d.f. $f_X(.)$, and we observe the fuzzy values $\widetilde{X}_1, \ldots, \widetilde{X}_n$. We want now to obtain a fuzzy density estimation with an extended empirical c.d.f. method.

**Definition 2** Let $\widetilde{X}_i[\alpha],\ i = 1, \ldots, n$ be the $\alpha$-cuts of the fuzzy numbers $\widetilde{X}_i,\ i = 1, \ldots, n$. The $\alpha$-cuts of fuzzy density estimation with empirical c.d.f. method are defined as

$$\begin{aligned}\widehat{f_n}(t)[\alpha] &= \frac{1}{2nh_n}\sum_{i=1}^{n}\left[\inf\left(I_{(t-h_n,t+h_n]}(y);\ y \in \widetilde{X}_i[\alpha]\right),\ \sup\left(I_{(t-h_n,t+h_n]}(y);\ y \in \widetilde{X}_i[\alpha]\right)\right]\\ &= \left[\frac{1}{2nh_n}\sum_{i=1}^{n}\inf\left(I_{(t-h_n,t+h_n]}(y);\ y \in \widetilde{X}_i[\alpha]\right),\right.\\ &\qquad \left.\frac{1}{2nh_n}\sum_{i=1}^{n}\sup\left(I_{(t-h_n,t+h_n]}(y);\ y \in \widetilde{X}_i[\alpha]\right)\right],\end{aligned}$$

where

$$I_{(t-h_n,t+h_n]}(y) = \left\{\begin{array}{l}1 \text{ for } t-h_n < y \le t+h_n\\ 0 \text{ otherwise}\end{array}\right\} \quad \forall\, y \in \Re.$$

**Theorem 1** *Consider the fuzzy empirical c.d.f. method in Definition 2.*

(a) *For every $Y_i \in \widetilde{\mathcal{X}}_i[\alpha]$, $i = 1, \ldots, n$, the expectation of $\hat{f}^L_\alpha$ (or $\hat{f}^U_\alpha$) converges to $f(t)$, if $h_n \to 0$ as $n \to \infty$.*

(b) *For every $Y_i \in \widetilde{\mathcal{X}}_i[\alpha]$, $i = 1, \ldots, n$, the variance of $\hat{f}^L_\alpha$ (or $\hat{f}^U_\alpha$) converges to 0, if $h_n \to 0$ and $nh_n \to \infty$.*

*Proof* Let $\widetilde{\mathcal{X}}[\alpha] = [\widetilde{\mathcal{X}}^L, \widetilde{\mathcal{X}}^U]$, then, for every $Y \in \widetilde{\mathcal{X}}[\alpha]$, we have

$$\begin{aligned}\inf\left(I_{(t-h_n,t+h_n]}(Y)\right) &= \begin{cases}1 & \text{for } t-h_n < \widetilde{\mathcal{X}}^L \le \widetilde{\mathcal{X}}^U \le t+h_n\\ 0 & \text{otherwise}\end{cases}\\ &= \left\{\begin{array}{ll}1 & \text{for } t-h_n < Y \le t+h_n,\\ 0 & \text{otherwise.}\end{array}\right\} \quad \forall\, Y \in \widetilde{\mathcal{X}}[\alpha],\end{aligned}$$

and

$$\begin{aligned}\sup\left(I_{(t-h_n,t+h_n]}(Y)\right) &= \begin{cases}0 & \text{for } \widetilde{\mathcal{X}}_U \le t-h_n \text{ or } \widetilde{\mathcal{X}}^L > t+h_n\\ 1 & \text{otherwise}\end{cases}\\ &= \left\{\begin{array}{ll}0 & \text{for } Y \le t-h_n \text{ or } Y > t+h_n,\\ 1 & \text{otherwise.}\end{array}\right\} \quad \forall\, Y \in \widetilde{\mathcal{X}}[\alpha].\end{aligned}$$

Hence, for every $Y \in \widetilde{\mathcal{X}}[\alpha]$, $\inf\left(I_{(t-h_n,t+h_n]}(Y)\right)$ and $\sup\left(I_{(t-h_n,t+h_n]}(Y)\right)$ have the Bernoulli distribution $Bernoulli(P(t-h_n < Y \le t+h_n))$.

(a) For every $Y_i \in \widetilde{\mathcal{X}}_i[\alpha]$, $i = 1, \ldots, n$, the expectations of $\hat{f}_\alpha^L$ and $\hat{f}_\alpha^U$ are calculated as

$$E(\hat{f}_\alpha^U) = E(\hat{f}_\alpha^L) = \frac{1}{2nh_n}\sum_{i=1}^{n} E\left[\inf\left(I_{(t-h_n,t+h_n]}(Y);\ Y \in \widetilde{\mathcal{X}}_i[\alpha]\right)\right]$$
$$= \frac{\int_{t-h_n}^{t+h_n} f(y)dy}{2h_n} \to f(t) \quad if\ h_n \to 0\ as\ n \to \infty.$$

(b) For every $Y_i \in \widetilde{\mathcal{X}}_i[\alpha]$, $i = 1, \ldots, n$, the variances of $\hat{f}_\alpha^L$ and $\hat{f}_\alpha^U$ are calculated as

$$Var(\hat{f}_\alpha^U) = Var(\hat{f}_\alpha^L) = \frac{1}{4n^2h_n^2}\sum_{i=1}^{n} Var\left[\inf\left(I_{(t-h_n,t+h_n]}(Y);\ Y \in \widetilde{\mathcal{X}}_i[\alpha]\right)\right]$$
$$= \frac{\left(\int_{t-h_n}^{t+h_n} f(y)dy\right)\left(1-\int_{t-h_n}^{t+h_n} f(y)dy\right)}{4nh_n^2}.$$

Using Hospital's rule, we obtain

$$\lim_{h_n\to 0,\ nh_n\to\infty} Var(\hat{f}_\alpha^L) = \lim_{h_n\to 0,\ nh_n\to\infty} \frac{\left(\int_{t-h_n}^{t+h_n} f(y)dy\right)\left(1-\int_{t-h_n}^{t+h_n} f(y)dy\right)}{4nh_n^2}$$
$$= \lim_{h_n\to 0,\ nh_n\to\infty} \frac{\left(f(t+h_n)+f(t-h_n)\right)\left(1-2\int_{t-h_n}^{t+h_n} f(y)dy\right)}{8nh_n}$$
$$= 0.$$

□

*Example 2* Consider the fuzzy data in Example 1. We want to construct a fuzzy density estimation based on the empirical c.d.f. method. Fig. 2 shows the fuzzy density estimation with $h_n = 0.5$ for some values of $t$ ($t = 0.5, 1, 1.5, 2, 2.5$). Some $\alpha$-cuts of the related results are shown in Table 2.

The fuzzy density estimation based on empirical c.d.f. method with $h_n = 0.5$ shows that it is approximately $\frac{7}{30}$ for $t = 0.5$, approximately $\frac{11}{30}$ for $t = 1$, approximately $\frac{13}{30}$ for $t = 1.5$, approximately $\frac{7}{15}$ for $t = 2$, and approximately $\frac{1}{3}$ for $t = 2.5$ (see Fig. 2). For example, the estimate of the density for the value $t = 2$, $\hat{f}_n(t = 2)$ is "approximately $\frac{7}{15}$", which is described quantitatively by a fuzzy number. Some $\alpha$-cuts of this fuzzy number are given in Table 2.

## 5 Fuzzy density estimation based on kernel method

The density estimations based on histogram and empirical c.d.f. methods are informative, but they suffer from two serious drawbacks: they are not smooth, and they are not

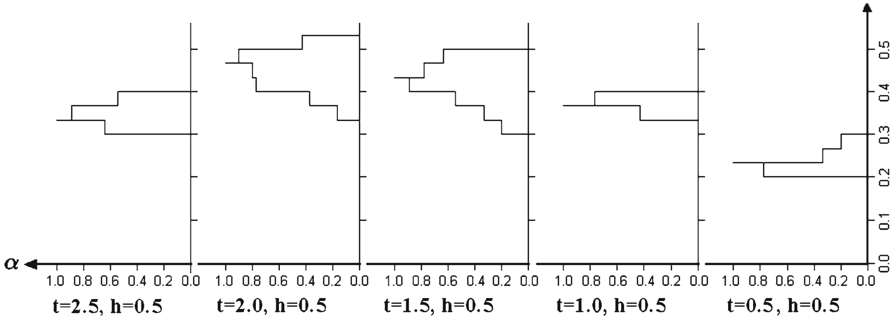


**Fig. 2** Fuzzy density estimation based on empirical c.d.f. method, with $h_n = 0.5$, in Example 2

**Table 2** Some $\alpha$-cuts of the fuzzy density estimation, based on empirical c.d.f. method, in Example 2

| | $t = 0.5$ $h_n = 0.5$ | $t = 1.0$ $h_n = 0.5$ | $t = 1.5$ $h_n = 0.5$ | $t = 2.0$ $h_n = 0.5$ | $t = 2.5$ $h_n = 0.5$ |
|---|---|---|---|---|---|
| $\alpha = 0.0$ | [0.2000, 0.3000] | [0.3333, 0.4000] | [0.3000, 0.5000] | [0.3333, 0.5667] | [0.3000, 0.4000] |
| $\alpha = 0.2$ | [0.2000, 0.3000] | [0.3333, 0.4000] | [0.3000, 0.5000] | [0.3667, 0.5333] | [0.3000, 0.4000] |
| $\alpha = 0.4$ | [0.2000, 0.2333] | [0.3333, 0.4000] | [0.3667, 0.5000] | [0.4000, 0.5333] | [0.3000, 0.4000] |
| $\alpha = 0.6$ | [0.2000, 0.2333] | [0.3667, 0.4000] | [0.4000, 0.5000] | [0.4000, 0.5000] | [0.3000, 0.3667] |
| $\alpha = 0.8$ | [0.2333, 0.2333] | [0.3667, 0.3667] | [0.4000, 0.4333] | [0.4667, 0.5000] | [0.3333, 0.3667] |
| $\alpha = 1.0$ | [0.2333, 0.2333] | [0.3667, 0.3667] | [0.4333, 0.4333] | [0.4667, 0.4667] | [0.3333, 0.3333] |

sensitive enough to local properties of $f(\cdot)$. It is easy to solve both of these problems. The histogram method uses Eq. (1) by dividing the line into bins, but a more sensible approach is to estimate the derivative separately at each point $t$. Replacing $F_X(t)$ by the empirical c.d.f. gives equation (2). This equation can be rewritten as

$$\hat{f}_n(t) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - x_i}{h_n}\right), \tag{3}$$

where

$$K(u) = \left\{ \begin{array}{ll} 0.5 & |u| \leq 1 \\ 0 & \text{otherwise} \end{array} \right\} \quad \forall\, u \in \Re.$$

The expression (3) is that of a *kernel density estimator* with kernel function $K(\cdot)$. Note that this kernel function is a uniform density function on $[-1, 1]$. Since the uniform density is discontinuous, we can use a smoother kernel function to lead to a smoother density estimation. The kernel function $K(.)$ is usually (but not necessarily) a symmetric p.d.f. itself, being positive valued, fulfilling the following conditions:

$$\begin{cases} \int_{-\infty}^{\infty} |K(y)|dy < \infty \\ \sup_y |K(y)| < \infty \\ \lim_{y\to\infty} |yK(y)| = 0 \end{cases}$$

For the influence of the shape of kernel functions and the bandwidth compare Alberts and Karunamuni (2003), and Wu et al. (2007). Some useful kernels are given in the following table.

| Kernel | $K(\cdot)$ | Support |
|---|---|---|
| Uniform | $K(y) = 0.5$ | $\|y\| \leq 1$ |
| Triangle | $K(y) = 1 - \|y\|$ | $\|y\| \leq 1$ |
| Quartic | $K(y) = \frac{15}{16}(1 - y^2)^2$ | $\|y\| \leq 1$ |
| Gaussian | $K(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ | $\|y\| < \infty$ |

*Remark 1* In contrast to the histogram estimator which requires specification of the origin and bandwidth, the kernel density estimate requires only specification of the bandwidth $h_n$. However, $h_n$ may be chosen as a function of the kernel function $K(\cdot)$, and the number of observations in the sample. For more details and a practical description of density estimation based on kernel method, see Silverman (1986), Wand and Jones (1995), Campos and Dorea (2001), Sheather (2004) and Ker and Ergün (2005).

Now, suppose that, we have i.i.d. FRVs $\widetilde{\mathcal{X}}_1, \ldots, \widetilde{\mathcal{X}}_n$ from a population with p.d.f. $f_X(.)$, and we observe the fuzzy values $\widetilde{X}_1, \ldots, \widetilde{X}_n$. We want to obtain a fuzzy density estimation based on kernel method.

**Definition 3** Let $\widetilde{X}_i[\alpha]$, $i = 1, \ldots, n$ be the $\alpha$-cuts of the fuzzy numbers $\widetilde{X}_i$, $i = 1, \ldots, n$. The $\alpha$-cuts of the fuzzy density estimation based on kernel method are defined as follows:

$$
\begin{aligned}
\hat{f}_n(t)[\alpha] &= \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - \widetilde{X}_i[\alpha]}{h_n}\right) \\
&= \frac{1}{nh_n} \sum_{i=1}^{n} \left[\inf\left(K\left(\frac{t-y}{h_n}\right);\ y \in \widetilde{X}_i[\alpha]\right), \sup\left(K\left(\frac{t-y}{h_n}\right);\ \ y \in \widetilde{X}_i[\alpha]\right)\right] \\
&= \left[\frac{1}{nh_n}\sum_{i=1}^{n} \inf\left(K\left(\frac{t-y}{h_n}\right);\ y \in \widetilde{X}_i[\alpha]\right), \frac{1}{nh_n}\sum_{i=1}^{n} \sup\left(K\left(\frac{t-y}{h_n}\right);\ y \in \widetilde{X}_i[\alpha]\right)\right].
\end{aligned}
$$

*Remark 2* The fuzzy density estimation based on empirical c.d.f. method is a special case of the fuzzy density estimation based on kernel method with $K(y) = 0.5$, $|y| \leq 1$. Note that

$$
\frac{1}{2} I_{(t-h_n, t+h_n]}(y) = \begin{cases} 0.5 & t - h_n < y \leq t + h_n, \\ 0 & \text{otherwise}, \end{cases} = \begin{cases} 0.5 & \left|\frac{t-y}{h_n}\right| \leq 1, \\ 0 & \text{otherwise}. \end{cases} = K\left(\frac{t-y}{h_n}\right).
$$

**Theorem 2** *Consider the fuzzy density estimation based on kernel method in Definition 3. Let $f(\cdot)$ be a bounded and continuous function at $t$.*

(a) *For every $Y_i \in \widetilde{\mathcal{X}}_i[\alpha]$, $i = 1, \ldots, n$, the bias of $\hat{f}_\alpha^L$ (or $\hat{f}_\alpha^U$) converges to 0, if $h_n \to 0$ as $n \to \infty$.*

(b) *For every* $Y_i \in \widetilde{\mathcal{X}}_i[\alpha]$, $i = 1, \ldots, n$, *the variance of* $\hat{f}_\alpha^L$ *(or* $\hat{f}_\alpha^U$*) converges to 0, if* $\int_{-\infty}^{\infty} \left(\inf\left(K(z);\ t - h_n z \in \widetilde{\mathcal{X}}_1[\alpha]\right)\right)^2 dz < \infty$, $h_n \to 0$, *and* $nh_n \to \infty$.

*Proof* (a) The bias of $\hat{f}_\alpha^L$ is calculated as (note that $\widetilde{\mathcal{X}}_i$'s are i.i.d)

$$Bias(\hat{f}_\alpha^L) = E(\hat{f}_\alpha^L) - f(t) = E\left(\frac{1}{nh_n}\sum_{i=1}^{n}\inf\left(K\left(\frac{t-Y}{h_n}\right);\ Y \in \widetilde{\mathcal{X}}_i[\alpha]\right)\right) - f(t)$$

$$= E\left(\frac{1}{h_n}\inf\left(K\left(\frac{t-Y}{h_n}\right);\ Y \in \widetilde{\mathcal{X}}_1[\alpha]\right)\right) - f(t)$$

$$= \frac{1}{h_n}\int_{-\infty}^{\infty}\inf\left(K\left(\frac{t-y}{h_n}\right);\ y \in \widetilde{\mathcal{X}}_1[\alpha]\right) f(y)dy - f(t)$$

$$= \int_{-\infty}^{\infty}\inf\left(K(z);\ t - h_n z \in \widetilde{\mathcal{X}}_1[\alpha]\right)[f(t-h_n z) - f(t)]dz.$$

If $f$ is bounded and continuous at $t$, then, based on the Lebesgue's dominated convergence theorem (Billingsley 1995, pp. 209), the bias of $\hat{f}_\alpha^L$ converges to 0 as follows:

$$\lim_{h_n\to 0} Bias(\hat{f}_\alpha^L) = \lim_{h_n\to 0}\int_{-\infty}^{\infty}\inf\left(K(z);\ t - h_n z \in \widetilde{\mathcal{X}}_1[\alpha]\right)[f(t-h_n z) - f(t)]dz$$

$$= \int_{-\infty}^{\infty}\lim_{h_n\to 0}\left[\inf\left(K(z);\ t - h_n z \in \widetilde{\mathcal{X}}_1[\alpha]\right)[f(t-h_n z) - f(t)]\right]dz$$

$$= 0.$$

(b) The variance of $\hat{f}_\alpha^L$ in the kernel density estimator is as

$$Var(\hat{f}_\alpha^L) = \frac{1}{nh_n^2}Var\left(\inf\left(K\left(\frac{t-Y}{h_n}\right);\ Y \in \widetilde{\mathcal{X}}_1[\alpha]\right)\right)$$

$$= \frac{1}{nh_n^2}\int_{-\infty}^{\infty}\left(\inf\left(K\left(\frac{t-y}{h_n}\right);\ y \in \widetilde{\mathcal{X}}_1[\alpha]\right)\right)^2 f(y)dy$$

$$-\frac{1}{n}\left[\frac{1}{h_n}\int_{-\infty}^{\infty}\left(\inf\left(K\left(\frac{t-y}{h_n}\right);\ y \in \widetilde{\mathcal{X}}_1[\alpha]\right)\right) f(y)dy\right]^2$$

$$= \frac{1}{nh_n}\int_{-\infty}^{\infty}\left(\inf\left(K(z);\ t - h_n z \in \widetilde{\mathcal{X}}_1[\alpha]\right)\right)^2 f(t-h_n z)dz + O\left(\frac{1}{n}\right)$$

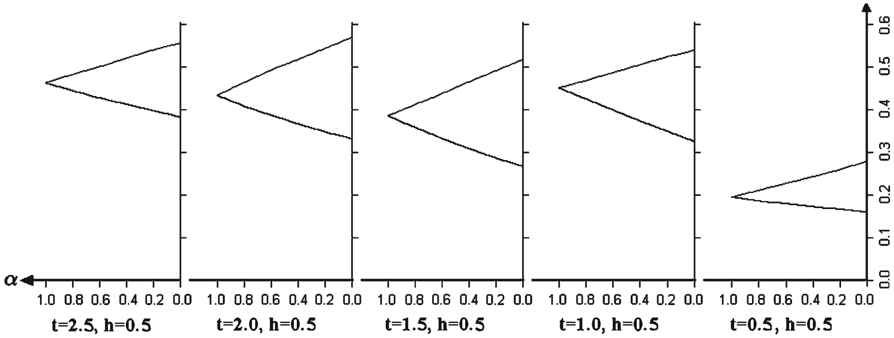


**Fig. 3** Fuzzy density estimation based on the triangle kernel function, with $h_n = 0.5$, in Example 3

**Table 3** Some $\alpha$-cuts of the fuzzy density estimation based on the triangle kernel function, with $h_n = 0.5$, in Example 3

| | $t = 0.5$ $h_n = 0.5$ | $t = 1.0$ $h_n = 0.5$ | $t = 1.5$ $h_n = 0.5$ | $t = 2.0$ $h_n = 0.5$ | $t = 2.5$ $h_n = 0.5$ |
|---|---|---|---|---|---|
| $\alpha = 0.0$ | [0.1600, 0.2787] | [0.3253, 0.5387] | [0.2667, 0.5173] | [0.3320, 0.5693] | [0.3827, 0.5560] |
| $\alpha = 0.2$ | [0.1664, 0.2592] | [0.3493, 0.5235] | [0.2861, 0.4912] | [0.3477, 0.5435] | [0.3976, 0.5397] |
| $\alpha = 0.4$ | [0.1728, 0.2427] | [0.3733, 0.5053] | [0.3085, 0.4651] | [0.3664, 0.5176] | [0.4125, 0.5205] |
| $\alpha = 0.6$ | [0.1792, 0.2267] | [0.3989, 0.4867] | [0.3323, 0.4373] | [0.3869, 0.4925] | [0.4275, 0.5003] |
| $\alpha = 0.8$ | [0.1859, 0.2107] | [0.4248, 0.4683] | [0.3581, 0.4115] | [0.4080, 0.4632] | [0.4448, 0.4821] |
| $\alpha = 1.0$ | [0.1947, 0.1947] | [0.4507, 0.4507] | [0.3853, 0.3853] | [0.4333, 0.4333] | [0.4627, 0.4627] |

$$= \frac{f(t)}{nh_n} \int_{-\infty}^{\infty} \left(\inf\left(K(z);\ t - h_n z \in \widetilde{\mathcal{X}}_1[\alpha]\right)\right)^2 dz + o\left(\frac{1}{nh_n}\right).$$

If $f$ is bounded and continuous at $t$ and $\int_{-\infty}^{\infty} \left(\inf\left(K(z);\ t - h_n z \in \widetilde{\mathcal{X}}_1[\alpha]\right)\right)^2 dz < \infty$, then, the variance of $\hat{f}_\alpha^L$ converges to 0 as $h_n \to 0$ and $nh_n \to \infty$. The bias and the variance of $\hat{f}_\alpha^U$ are calculated analogous. □

*Example 3* Consider the fuzzy data in Example 1. Now, we want to estimate the density based on the extended kernel method. Fig. 3 shows the fuzzy density estimation with the kernel function $K(y) = (1 - |y|)I_{[-1,1]}$, with $h_n = 0.5$, for some values of $t$. Some $\alpha$-cuts of the obtained fuzzy density estimation are given in Table 3.

Therefore, the fuzzy density estimation, based on $K(y) = (1 - |y|)I_{[-1,1]}$ with $h_n = 0.5$, is obtained as approximately 0.1947 for $t = 0.5$, approximately 0.4507 for $t = 1$, approximately 0.3853 for $t = 1.5$, approximately 0.4333 for $t = 2$, and approximately 0.4627 for $t = 2.5$ (see Fig. 3).

Also, the estimate of the p.d.f. based on the Gaussian kernel function $K(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ is calculated. Fig. 4 shows the fuzzy density estimation based on the

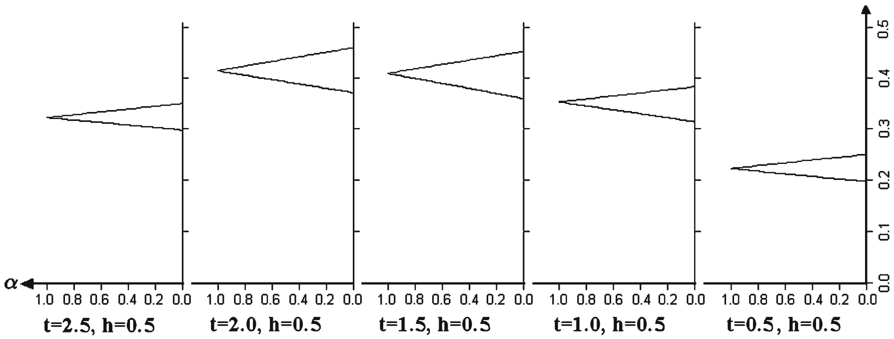


**Fig. 4** Fuzzy density estimation based on the Gaussian kernel function, with $h_n = 0.5$, in Example 3

**Table 4** Some $\alpha$-cuts of the fuzzy density estimation based on the Gaussian kernel function, with $h_n = 0.5$, in Example 3

| | $t = 0.5$ $h_n = 0.5$ | $t = 1.0$ $h_n = 0.5$ | $t = 1.5$ $h_n = 0.5$ | $t = 2.0$ $h_n = 0.5$ | $t = 2.5$ $h_n = 0.5$ |
|---|---|---|---|---|---|
| $\alpha = 0.0$ | [0.1977, 0.2503] | [0.3142, 0.3821] | [0.3593, 0.4510] | [0.3707, 0.4584] | [0.2980, 0.3498] |
| $\alpha = 0.2$ | [0.2025, 0.2449] | [0.3221, 0.3762] | [0.3692, 0.4426] | [0.3792, 0.4495] | [0.3030, 0.3443] |
| $\alpha = 0.4$ | [0.2075, 0.2395] | [0.3299, 0.3703] | [0.3792, 0.4341] | [0.3878, 0.4406] | [0.3079, 0.3388] |
| $\alpha = 0.6$ | [0.2125, 0.2340] | [0.3375, 0.3644] | [0.3890, 0.4256] | [0.3963, 0.4316] | [0.3128, 0.3333] |
| $\alpha = 0.8$ | [0.2177, 0.2285] | [0.3451, 0.3584] | [0.3988, 0.4171] | [0.4049, 0.4225] | [0.3175, 0.3277] |
| $\alpha = 1.0$ | [0.2230, 0.2230] | [0.3525, 0.3525] | [0.4085, 0.4085] | [0.4134, 0.4134] | [0.3221, 0.3221] |

Gaussian kernel function for some values of $t$. For example, the estimate of the density for the value $t = 1.5$, $\hat{f}_n(t = 1.5)$ is "approximately 0.4085". This result is described quantitatively by a fuzzy number. Some $\alpha$-cuts of this fuzzy number are given in Table 4.

## 6 A practical example

In this section, we apply our proposed methods in a real world problem, considering a well-known data set.

*Example 4* The following data (the centers of fuzzy numbers) are the amounts of winter snowfall (in inches) at Buffalo, New York, for each of 63 winters from 1910/1911 to 1972/1973. These data have been considered by several authors, for example, Parzen (1962) and Silverman (1986). But, imprecision and vagueness in measurements (especially in natural events, which occur frequently in practice) may yield fuzzy quantities as collected data. Assume that the amounts of winter snowfall are reported as fuzzy numbers in the table. In fact, imprecision is formulated by symmetric triangular fuzzy numbers, with $s_i = 0.05a_i$, $i = 1, 2, \ldots, 63$.

| No. | Data | No. | Data | No. | Data | No. | Data |
|---|---|---|---|---|---|---|---|
| 1 | $(126.4, 6.3)_T$ | 17 | $(79.6, 4.0)_T$ | 33 | $(89.6, 4.5)_T$ | 49 | $(124.7, 6.2)_T$ |
| 2 | $(82.4, 4.1)_T$ | 18 | $(83.6, 4.2)_T$ | 34 | $(85.5, 4.3)_T$ | 50 | $(114.5, 5.7)_T$ |
| 3 | $(78.1, 3.9)_T$ | 19 | $(80.7, 4.0)_T$ | 35 | $(58.0, 2.9)_T$ | 51 | $(115.6, 5.8)_T$ |
| 4 | $(51.1, 2.6)_T$ | 20 | $(60.3, 3.0)_T$ | 36 | $(120.7, 6.0)_T$ | 52 | $(102.4, 5.1)_T$ |
| 5 | $(90.9, 4.5)_T$ | 21 | $(79.0, 4.0)_T$ | 37 | $(110.5, 5.5)_T$ | 53 | $(101.4, 5.1)_T$ |
| 6 | $(76.2, 3.8)_T$ | 22 | $(74.4, 3.7)_T$ | 38 | $(65.4, 3.3)_T$ | 54 | $(89.8, 4.5)_T$ |
| 7 | $(104.5, 5.2)_T$ | 23 | $(49.6, 2.5)_T$ | 39 | $(39.9, 2.0)_T$ | 55 | $(71.5, 3.6)_T$ |
| 8 | $(87.4, 4.4)_T$ | 24 | $(54.7, 2.7)_T$ | 40 | $(40.1, 2.0)_T$ | 56 | $(70.9, 3.5)_T$ |
| 9 | $(110.5, 5.5)_T$ | 25 | $(71.8, 3.6)_T$ | 41 | $(88.7, 4.4)_T$ | 57 | $(98.3, 4.9)_T$ |
| 10 | $(25.0, 1.2)_T$ | 26 | $(49.1, 2.5)_T$ | 42 | $(71.4, 3.6)_T$ | 58 | $(55.5, 2.8)_T$ |
| 11 | $(69.3, 3.5)_T$ | 27 | $(103.9, 5.2)_T$ | 43 | $(83.0, 4.2)_T$ | 59 | $(66.1, 3.3)_T$ |
| 12 | $(53.5, 2.7)_T$ | 28 | $(51.6, 2.6)_T$ | 44 | $(55.9, 2.8)_T$ | 60 | $(78.4, 3.9)_T$ |
| 13 | $(39.8, 2.0)_T$ | 29 | $(82.4, 4.1)_T$ | 45 | $(89.9, 4.5)_T$ | 61 | $(120.5, 6.0)_T$ |
| 14 | $(63.6, 3.2)_T$ | 30 | $(83.6, 4.2)_T$ | 46 | $(84.8, 4.2)_T$ | 62 | $(97.0, 4.9)_T$ |
| 15 | $(46.7, 2.3)_T$ | 31 | $(77.8, 3.9)_T$ | 47 | $(105.2, 5.3)_T$ | 63 | $(110.0, 5.5)_T$ |
| 16 | $(72.9, 3.6)_T$ | 32 | $(79.3, 4.0)_T$ | 48 | $(113.7, 5.7)_T$ | | |

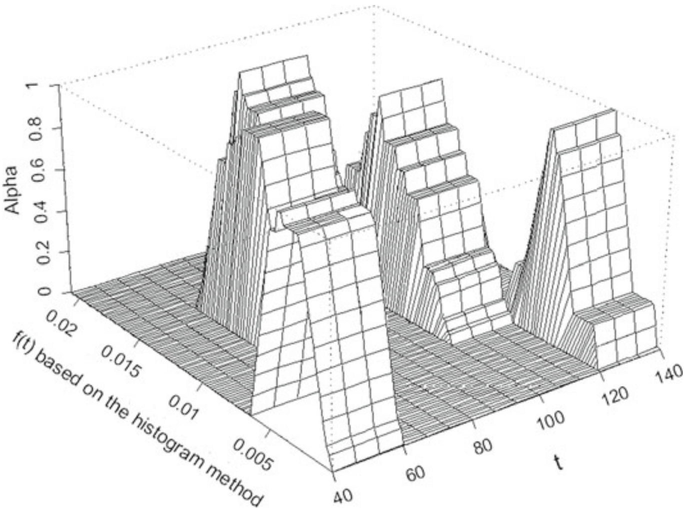


**Fig. 5** Fuzzy density estimation based on histogram method, in Example 4

Now, we estimate the p.d.f. based on these fuzzy data, using the proposed methods which are introduced in the sections before.

**(I) Fuzzy density estimation based on histogram method:**
The 3-dimensional curve of the fuzzy density estimation based on the histogram method is shown in Fig. 5 for $t \in [40, 60), [60, 80), [80, 100), [100, 120), [120, 140]$. Also, the fuzzy density estimation for $t = 66, 90, 116$ is shown in Fig. 6 (see also, 3-D curve in Fig. 5). Some $\alpha$-cuts of the fuzzy density estimation, for $t = 66, 90, 116$, are given in Table 5.

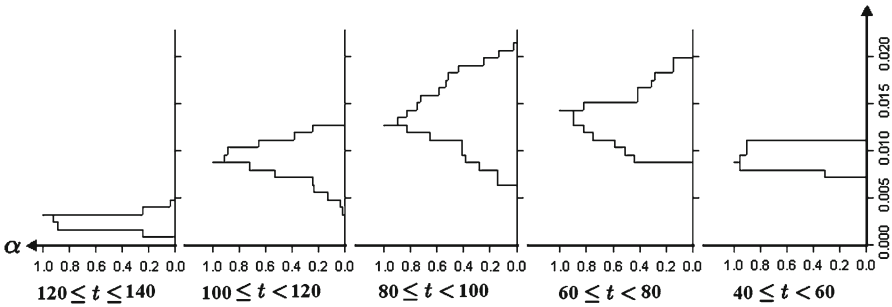


**Fig. 6** Fuzzy density estimation based on histogram method, for $t = 66, 90, 116$, in Example 4

**Table 5** Some $\alpha$-cuts of the fuzzy density estimation, based on histogram method, in Example 4

| | $40 \leq t < 60$ | $60 \leq t < 80$ | $80 \leq t < 100$ | $100 \leq t < 120$ | $120 \leq t < 140$ |
|---|---|---|---|---|---|
| $\alpha = 0.0$ | [0.0071, 0.0079] | [0.0063, 0.0032] | [0.0008, 0.0111] | [0.0198, 0.0222] | [0.0127, 0.0048] |
| $\alpha = 0.2$ | [0.0071, 0.0087] | [0.0079, 0.0056] | [0.0008, 0.0111] | [0.0183, 0.0198] | [0.0127, 0.0040] |
| $\alpha = 0.4$ | [0.0079, 0.0087] | [0.0095, 0.0071] | [0.0016, 0.0111] | [0.0167, 0.0190] | [0.0111, 0.0032] |
| $\alpha = 0.6$ | [0.0079, 0.0111] | [0.0111, 0.0079] | [0.0016, 0.0111] | [0.0151, 0.0159] | [0.0111, 0.0032] |
| $\alpha = 0.8$ | [0.0079, 0.0119] | [0.0119, 0.0087] | [0.0016, 0.0111] | [0.0151, 0.0143] | [0.0103, 0.0032] |
| $\alpha = 1.0$ | [0.0087, 0.0143] | [0.0127, 0.0087] | [0.0032, 0.0087] | [0.0143, 0.0127] | [0.0087, 0.0032] |

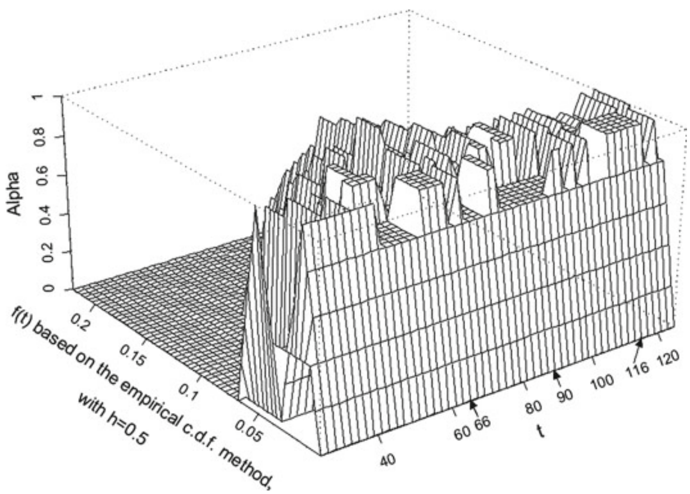


**Fig. 7** Fuzzy density estimation based on empirical c.d.f. method, with $h_n = 0.5$, in Example 4

**(II) Fuzzy density estimation based on empirical c.d.f. method:**
Figure 7 show 3-dimensional curve of the extended density estimation based on empirical c.d.f. method with $h_n = 0.5$ for $t = 25, 27, 29, \ldots, 125$.

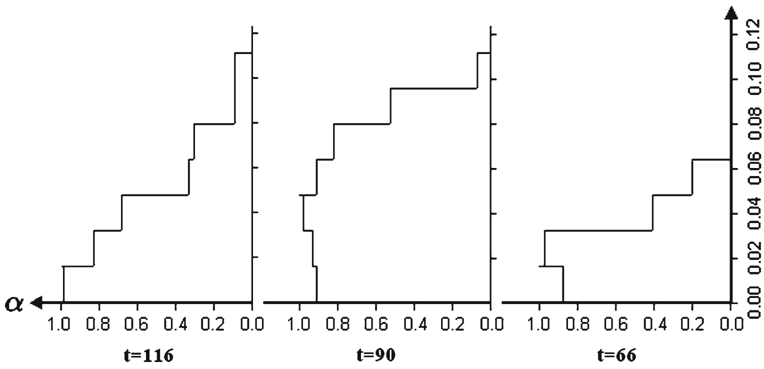


**Fig. 8** Fuzzy density estimation based on empirical c.d.f. method, with $h_n = 0.5$, in Example 4

**Table 6** Some $\alpha$-cuts of the fuzzy density estimation based on empirical c.d.f. method, with $h_n = 0.5$, in Example 4

| | $t = 66$ | $t = 90$ | $t = 116$ |
|---|---|---|---|
| $\alpha = 0.0$ | [0.0000, 0.0635] | [0.0000, 0.1111] | [0.0000, 0.1111] |
| $\alpha = 0.2$ | [0.0000, 0.0635] | [0.0000, 0.0952] | [0.0000, 0.0794] |
| $\alpha = 0.4$ | [0.0000, 0.0476] | [0.0000, 0.0952] | [0.0000, 0.0476] |
| $\alpha = 0.6$ | [0.0000, 0.0317] | [0.0000, 0.0794] | [0.0000, 0.0476] |
| $\alpha = 0.8$ | [0.0000, 0.0317] | [0.0000, 0.0794] | [0.0000, 0.0317] |
| $\alpha = 1.0$ | [0.0159, 0.0159] | [0.0476, 0.0476] | [0.0159, 0.0159] |

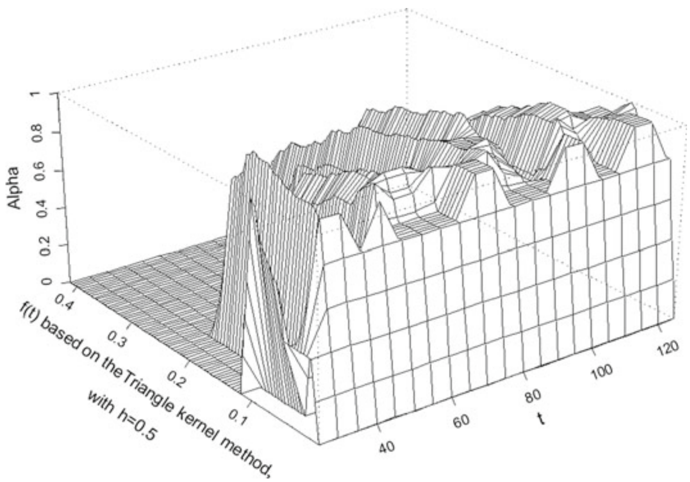


**Fig. 9** Fuzzy density estimation based on the triangle kernel function, with $h_n = 0.5$, in Example 4

Also, Fig. 8 shows fuzzy density estimation based on empirical c.d.f. method, with $h_n = 0.5$, for $t = 66, 90, 116$ (see also, 3-D curve in Fig. 7). Some $\alpha$-cuts of the fuzzy density estimation, for $t = 66, 90, 116$, are given in Table 6.

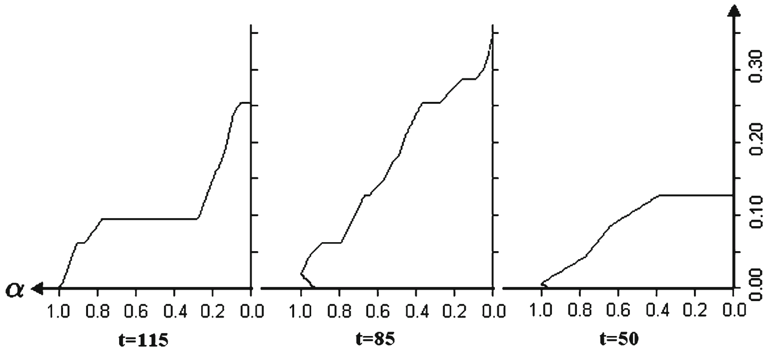


**Fig. 10** Fuzzy density estimation based on triangle kernel, with $h_n = 0.5$, for $t = 50, 85, 115$, in Example 4

**Table 7** Some $\alpha$-cuts of the fuzzy density estimation based on the triangle kernel function, with $h_n = 0.5$, in Example 4

| | $t = 66$ | $t = 90$ | $t = 116$ |
|---|---|---|---|
| $\alpha = 0.0$ | [0.0000, 0.1270] | [0.0000, 0.3429] | [0.0000, 0.2540] |
| $\alpha = 0.2$ | [0.0000, 0.1270] | [0.0000, 0.2743] | [0.0000, 0.1460] |
| $\alpha = 0.4$ | [0.0000, 0.1244] | [0.0000, 0.2362] | [0.0000, 0.0952] |
| $\alpha = 0.6$ | [0.0000, 0.0914] | [0.0000, 0.1384] | [0.0000, 0.0952] |
| $\alpha = 0.8$ | [0.0000, 0.0381] | [0.0000, 0.0635] | [0.0000, 0.0851] |
| $\alpha = 1.0$ | [0.0063, 0.0063] | [0.0190, 0.0190] | [0.0000, 0.0000] |

**(III) Fuzzy density estimation based on kernel method:**
Figure 9 shows the 3-dimensional curve of fuzzy density estimation based on the kernel function $K(y) = (1-|y|)I_{[-1,1]}$, with $h_n = 0.5$, for $t = 25, 30, \ldots, 125$. Also, Fig. 10 shows the fuzzy density estimation based on the triangle kernel function, with $h_n = 0.5$, for $t = 50, 85, 115$ (see also, 3-D curve in Fig. 9). Some $\alpha$-cuts of the fuzzy density estimation, for $t = 50, 85, 115$, are given in Table 7. By comparing the results of the fuzzy density estimation based on three methods, it is obvious that the extended kernel method is yielding a smoother estimation than the other ones, as shown in Figs. 5, 6, 7, 8, 9, and 10.

*Remark 3* It should be mentioned that the proposed methods specialize to the classical ones, when the data available are crisp. For instance, in Example 4, if we use the classical approaches for the crisp data (i.e. for the centers of the fuzzy numbers), then we obtain the same density estimations as we obtained for the one-point level set $\alpha = 1$.

## 7 Conclusion

In this paper, some methods for estimating p.d.f.s are extended when the available data are fuzzy. These extended methods (histogram, empirical c.d.f., and kernel methods) calculate $\alpha$-cuts of a fuzzy density estimation based on the $\alpha$-cuts of fuzzy data.

The results show that among three extended methods, the extended kernel method yields smooth density estimators, which, as in the classical case, is a considerable advantage.

**Acknowledgments** The authors wish to express their thanks to the referees for valuable comments which improved the paper. The third author is partially supported by Fuzzy Systems and Applications Center of Excellence at Shahid Bahonar University of Kerman, Iran.

## References

Alberts T, Karunamuni RJ (2003) A semiparametric method of boundary correction for kernel density estimation. Stat Probab Lett 61:287–298

Billingsley P (1995) Probability and measure, 3rd edn. Wiley, New York

Campos VSM, Dorea CCY (2001) Kernel density estimation: the general case. Stat Probab Lett 55:173–180

Cheng K, Chu C (2004) Semiparametric density estimation under a two-sample density ratio model. Bernoulli 10(4):583–604

Cheng PE (1994) Nonparametric estimation of mean functionals with data missing at random. J Am Stat Assoc 89:81–87

Cheng PE, Chu CK (1996) Kernel estimation of distribution functions and quantiles with missing data. Stat Sinica 6:63–78

Devroye L, Györfi L (1985) Nonparametric density estimation. Wiley, New York

Fokianos K (2004) Merging information for semiparametric density estimation. J R Stat Soc Series B (Stat Methodol) 66(4):941–958

Gil MA (2004) Fuzzy random variables: development and state of the art. In: Klement EP, Pap E (eds) Mathematics of fuzzy systems, linz seminar on fuzzy set theory. Linz, Austria, pp 11–15

Hazelton ML (2000) Marginal density estimation from incomplete bivariate data. Stat Probab Lett 47:75–84

Jones MC (1991) Kernel density estimation for length biased data. Biometrika 78:511–519

Ker AP, Ergün AT (2005) Empirical Bayes nonparametric kernel density estimation. Stat Probab Lett 75:315–324

Keziou A, Leoni-Aubin S (2005) Test of homogeneity in semiparametric two-sample density ratio models. C R Acad Sci Paris Ser I Math 340(12):905–910

Keziou A, Leoni-Aubin S (2007) On empirical likelihood for semiparametric two-sample density ratio models. J Stat Plan Inference 138:915–928

Klement EP, Puri LM, Ralescu DA (1986) Limit theorems for fuzzy random variables. Proc R Soc Lond 407:171–182

Klir GJ, Yuan B (1995) Fuzzy sets and fuzzy logic, theory and applications. Prentic-Hall, Englewood Cliffs, NJ

Kruse R, Meyer KD (1987) Statistics with vague data. Reidel Publishing Company, Dordrecht, Netherlands

Lee YK, Choi H, Park BU, Yu KS (2004) Local likelihood density estimation on random fields. Stat Probab Lett 68:347–357

Li S, Ogura Y (2006) Strong laws of large numbers for independent fuzzy set-valued random variables. Fuzzy Sets Syst 157:2569–2578

Loquin K, Strauss O (2008) Histogram density estimators based upon a fuzzy partition. Stat Probab Lett 78:1863–1868

Owen AB (2001) Empirical likelihood. Chapman & Hall/CRC, London

Parzen E (1962) On estimation of a probability density function and mode. Ann Math Stat 33:1065–1076

Prasaka Rao BLS (1983) Nonparametric functional estimation. Academic Press, New York

Puri ML, Ralescu DA (1986) Fuzzy random variables. J Math Anal Appl 114:409–422

Qin J (1998) Inferences for case-control and semiparametric two-sample density ratio models. Biometrika 85(3):619–630

Qin J, Zhang B (2005) Density estimation under a two-sample semiparametric model. Nonparametric Stat 17(6):665–683

Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. Ann Math Stat 27:642–669

Rosenblatt M (1971) Curve estimates. Ann Math Stat 42:1815–1842

Shao J (2003) Mathematical statistics, 2nd edn. Springer-Verlag, New York
Sheather SJ (2004) Density estimation. Stat Sci 19:588–597
Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall, New York
Simonoff J (1996) Smoothing methods in statistics. Springer, New York
Taheri SM (2003) Trends in fuzzy statistics. Austrian J Stat 32:239–257
Trutschnig W (2008) A strong consistency result for fuzzy relative frequencies interpreted as estimator for the fuzzy-valued probability. Fuzzy Sets Syst 159:259–269
Viertl R (1996) Statistical methods for non-precise data. CRC Press, Boca Raton
Viertl R (2006) Univariate statistical analysis with fuzzy data. Comput Stat Data Anal 51:133–147
Viertl R, Hareter D (2006) Beschreibung und Analyse unscharfer Information: Statistische Methoden für unscharfe Daten. Springer, Wien
Wand MP, Jones MC (1995) Kernel smoothing. Chapman & Hall, London
Wu TJ, Chen ChF, Chen HY (2007) A variable bandwidth selector in multivariate kernel density estimation. Stat Probab Lett 77:462–467