Maximum likelihood estimation from fuzzy data using the EM algorithm

Thierry Denœux
UMR CNRS 6599 Heudiasyc
Université de Technologie de Compiègne
BP 20529 - F-60205 Compiègne cedex - France
Thierry.Denoeux@hds.utc.fr

May 26, 2011

Abstract

A method is proposed for estimating the parameters in a parametric statistical model when the observations are fuzzy and are assumed to be related to underlying crisp realizations of a random sample. This method is based on maximizing the observed-data likelihood defined as the probability of the fuzzy data. It is shown that the EM algorithm may be used for that purpose, which makes it possible to solve a wide range of statistical problems involving fuzzy data. This approach, called the Fuzzy EM (FEM) method, is illustrated using three classical problems: normal mean and variance estimation from a fuzzy sample, multiple linear regression with crisp inputs and fuzzy outputs, and univariate finite normal mixture estimation from fuzzy data.

Keywords: Statistics, fuzzy data analysis, estimation, maximum likelihood principle, regression, mixture models.

1 Introduction

Recent years have seen a surge of interest in extending statistical inferential procedures to fuzzy data [5]. Basically, two views of fuzzy data have been proposed [13].

The first approach assumes the data to be intrinsically fuzzy and uses the mathematical formalism of fuzzy random variables, defined as mappings from a probability space to the set of fuzzy subsets of $\mathbb{R}$ or $\mathbb{R}^n$ [17, 24, 25, 29], with certain measurability properties. This model fits well with a *physical interpretation* of fuzzy data, in which a fuzzy datum is regarded as an existing object, not necessarily connected to any underlying precise variable [13]. Recent work on applying this approach to estimation and hypothesis testing problems may be found, for example, in [2, 19, 21].

The second approach is based on an *epistemic interpretation* of fuzzy data, which are assumed to "imperfectly specify a value that is existing and precise, but not measurable with exactitude under the given observation conditions" [13, page 316]. In this model, a fuzzy datum is thus seen as a possibility distribution associated to a precise realization of a random variable that has been only partially observed. This viewpoint will be adopted in this paper.

Under the latter interpretation of fuzzy data, a conceptually simple approach is to "fuzzify" standard statistical computations using Zadeh's extension principle [34]. This approach makes it possible to compute fuzzy versions of classical statistics, such as the mean, variance or correlation coefficients [8, 33], and even to extend significance tests using fuzzy p-values [8, 32]. However, it usually implies solving complex nonlinear optimization problems, except in very simple cases. Furthermore, fuzzy estimates are often too complex objects to be easily handled and interpreted by end-users.

When a parametric statistic model is postulated, an alternative approach to parameter estimation based on fuzzy data relies on the extension of the likelihood function using Zadeh's definition of the probability of a fuzzy event [35]. A maximum likelihood estimate (MLE) of the parameter of interest may then be defined as a crisp value maximizing the probability of observing the fuzzy data. This model was initially suggested in [28]. In the 1980's, it was studied by Gil and her colleagues in conjunction with fuzzy information systems [14, 15, 16]. These authors showed that the main properties of the MLE (equivariance under one-to-one transformations, asymptotic normality and asymptotic efficiency) are preserved in the fuzzy case under very general conditions [13, pages 327–333].

Although conceptually appealing and well-founded theoretically, the maximum likelihood (ML) approach to statistical inference from fuzzy data has not been widely used, essentially because of the difficulty of putting it to work in realistic statistical problems. As noted by Gebhardt et al. [13, page 329]: "In spite of these valuable properties and although the application of the maximum likelihood principle to some examples is very simple, to determine solutions in most situations becomes too complex".

In this paper, we propose to solve the above problem using the Expectation-Maximization (EM) algorithm. This algorithm, formalized by Dempster et al. [6], provides a very general mechanism for computing MLEs from incomplete data. This method has been successfully applied to a wide range of problems involving partially observed data such as inference from censored or grouped data, finite mixture models, factor analysis, etc [26]. As will be shown, the EM algorithm can be adapted to handle

estimation problems involving fuzzy data, viewed as imperfectly specifying underlying crisp values. As will be shown, this approach makes the ML principle practically applicable to solve a wide range of statistical problems involving fuzzy data.

The rest of this paper is organized as follows. Section 2 first presents in greater detail the problem addressed in this paper, trying to clarify the considered semantics of fuzzy data. Section 3 recalls the EM algorithm and demonstrates its application in the case of fuzzy data. Section 4 then describes three applications: normal mean and variance estimation of a normal distribution, multiple regression with crisp input and fuzzy output data, and univariate normal mixture estimation. Finally, Section 5 concludes the paper.

2 Problem description

The problem addressed in this paper may be described as follows. We assume the existence of a random vector $\mathbf{X}$, referred to as the *complete data* vector, taking values in a sample space $\mathcal{X}$ and describing the result of a random experiment. The probability density function (p.d.f.) of $\mathbf{X}$ is denoted by $g(\mathbf{x}; \Psi)$, where $\mathbf{\Psi} = (\Psi_1, \ldots, \Psi_d)'$ is a vector of unknown parameters with parameter space $\mathbf{\Omega}$. Although $\mathbf{X}$ will be generally assumed to be a continuous random vector (unless otherwise specified), $g_(\mathbf{x}; \Psi)$ can still be viewed as a p.d.f. in the case where $\mathbf{X}$ is discrete by the adoption of the counting measure.

Let $\mathbf{x}$ be a realization of $\mathbf{X}$. If $\mathbf{x}$ was known exactly, we could compute the maximum likelihood estimate (MLE) of $\mathbf{\Psi}$ as any value maximizing the complete-data likelihood function

$$L(\mathbf{\Psi}; \mathbf{x}) = g(\mathbf{x}; \mathbf{\Psi}). \tag{1}$$

In this paper, we consider the more difficult problem where $\mathbf{x}$ is not observed precisely, and only *partial information* about $\mathbf{x}$ is available in the form of a fuzzy subset $\widetilde{\mathbf{x}}$ of $\mathcal{X}$, with Borel measurable membership function $\mu_{\widetilde{\mathbf{x}}} : \mathcal{X} \to [0, 1]$.

Before tackling the problem of parameter estimation from such data, we have to examine carefully the nature of the partial information about $\mathbf{x}$, in relation to the underlying random experiment. This issue will be addressed in Subsection 2.1. A generalized likelihood function will then be introduced and discussed in Subsection 2.2. Finally, another model, which leads formally to the same likelihood function with a completely different interpretation, will be discussed in Subsection 2.3.

2.1 Interpretation of fuzzy data

In the model considered here, the fuzzy observation $\widetilde{\mathbf{x}}$ will be understood as encoding the observer's *partial knowledge* about the realization $\mathbf{x}$ of random vector $\mathbf{X}$. In this setting, the membership function $\mu_{\widetilde{\mathbf{x}}}$ is seen as a possibility distribution [36, 11] interpreted as a soft constraint on the unknown quantity $\mathbf{x}$. The fuzzy set $\widetilde{\mathbf{x}}$ can be considered to be generated by a two-step process:

1. A realization $\mathbf{x}$ is drawn from $\mathbf{X}$;
2. The observer encodes his/her partial knowledge of $\mathbf{x}$ in the form of a possibility distribution $\mu_{\widetilde{\mathbf{x}}}$.

It must be stressed that, in this model, only step 1 is considered to be a random experiment. Step 2 implies gathering information about $\mathbf{x}$ and modeling this information as a possibility distribution. It is not assumed that this process may be repeated indefinitely in the same experimental conditions, i.e., it is not considered as a random experiment. Consequently, $\widetilde{\mathbf{x}}$ is not considered as a realization of a fuzzy random variable in this paper. It must also be recognized that two kinds of uncertainty are involved in this setting:

- Step 1 is associated with *aleatory uncertainty*, which is due to the random nature of the data generation process; this uncertainty cannot be reduced before the experiment, and disappears after the experiment has been performed.
- Step 2 is associated by *epistemic uncertainty*, which is related to the observer's state of knowledge. This uncertainty can sometimes be reduced by gathering additional information about $\mathbf{x}$.

EXAMPLE 1 Assume that n skull fragments have been found in some archaeological site, and we are interested in the volumes of these skulls. The unknown volume x_i of skull i may be regarded as a realization of a random variable X_i induced by random sampling from a total population of skulls (corresponding to a certain region and period of interest). The complete data $\mathbf{x} = (x_1, \ldots, x_n)$ is thus a realization from a random vector $\mathbf{X} = (X_1, \ldots, X_n)$. As only fragments are available, archaeologists have to use various indirect methods and hypotheses to assess the volume of each skull. Assume that, following a procedure reported, e.g., in [8] and [20], two intervals are determined for each skull i:

- an interval $[a_i, d_i]$ certainly containing x_i;
- an interval $[b_i, c_i]$ containing highly plausible values for x_i.

This information may be encoded as a trapezoidal fuzzy number $\widetilde{x}_i = (a_i, b_i, c_i, d_i)$ with support $[a_i, d_i]$ and core $[b_i, c_i]$, interpreted as a possibility distribution constraining the unknown value x_i. Information about $\mathbf{x}$ may be represented by the joint possibility distribution

$$\mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) = \mu_{\widetilde{x}_1}(x_1) \times \cdots \times \mu_{\widetilde{x}_1}(x_n). \tag{2}$$

The epistemic uncertainty modeled by this possibility distribution might be reduced by carrying out further analysis of the skull fragments. This uncertainty is clearly of a different nature than the aleatory uncertainty involved in the sampling process. □

We may note that, in the special case where the observed data $\widetilde{\mathbf{x}}$ is crisp, it corresponds to some form of *censored data* as frequently encountered in statistics. Usually, censored data are handled by postulating some stochastic censoring mechanism, and maximizing the observed data likelihood using, e.g., the EM algorithm. As we will see, our approach is computationally very similar, although no random censoring mechanism is assumed.

The relationship between the unknown crisp realization $\mathbf{x}$ and the fuzzy observation $\widetilde{\mathbf{x}}$ can be described by considering the random set view of fuzzy sets (see, e.g., [10, page 46]). In this view, $\widetilde{\mathbf{x}}$ is seen as being induced by a uniform probability measure on $[0, 1]$ and the set-valued mapping $\alpha \rightarrow {}^{\alpha}\widetilde{\mathbf{x}}$, where ${}^{\alpha}\widetilde{\mathbf{x}}$ denotes the α-cut of $\widetilde{\mathbf{x}}$. If

the observer knew that $\alpha = \alpha_0$, then he/she would assert that $\mathbf{x} \in {}^{\alpha_0}\widetilde{\mathbf{x}}$. However, the observer's knowledge of α is uncertain and is described by the uniform probability distribution. In this model, $\mu_{\widetilde{\mathbf{x}}}(\mathbf{u})$ is the observer's *subjective* probability that the available evidence can be interpreted as specifying an interval that contains $\mathbf{u}$. We should stress here that, in the model adopted in this paper, the known subjective probability distribution on α and the unknown objective probability distribution on $\mathbf{X}$ are assumed to be of different natures; as such they will be treated in completely different ways in the approach described below.

2.2 Generalized likelihood function

Once $\widetilde{\mathbf{x}}$ is given, and assuming its membership function to be Borel measurable, we can compute its probability according to Zadeh's definition of the probability of a fuzzy event [35] (see Appendix A.1). By analogy with (1), the observed-data likelihood can then be defined as:

$$L(\boldsymbol{\Psi}; \widetilde{\mathbf{x}}) = P(\widetilde{\mathbf{x}}; \boldsymbol{\Psi}) = \int_{\mathcal{X}} \mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) g(\mathbf{x}; \boldsymbol{\Psi}) d\mathbf{x}. \tag{3}$$

To understand the meaning of $P(\widetilde{\mathbf{x}}; \boldsymbol{\Psi})$ and, consequently, of $L(\boldsymbol{\Psi}; \widetilde{\mathbf{x}})$, we can remark[1], as done by Höhle [23], that $P(\widetilde{\mathbf{x}}; \boldsymbol{\Psi})$ can also be written as:

$$P(\widetilde{\mathbf{x}}; \boldsymbol{\Psi}) = \int_0^1 P({}^{\alpha}\widetilde{x}; \boldsymbol{\Psi}) d\alpha,$$

where ${}^{\alpha}\widetilde{\mathbf{x}}$ denotes the α-cut of $\widetilde{\mathbf{x}}$ (see also [10, page 52]). Using the random set view of fuzzy sets outlined in the previous section, $P(\widetilde{\mathbf{x}}; \boldsymbol{\Psi})$ can thus be seen as the average value of $P({}^{\alpha}\widetilde{\mathbf{x}}; \boldsymbol{\Psi})$ over the random set underlying $\widetilde{\mathbf{x}}$.

The above remark can be reformulated in terms of likelihoods. If we knew that $\mathbf{x} \in {}^{\alpha}\widetilde{\mathbf{x}}$, then the likelihood function would be:

$$L(\boldsymbol{\Psi}; {}^{\alpha}\widetilde{\mathbf{x}}) = P({}^{\alpha}\widetilde{\mathbf{x}}; \boldsymbol{\Psi}).$$

Assuming α to be uncertain, and our belief on α to be described by a uniform probability distribution on $[0, 1]$, the likelihood averaged over all values of α is

$$\int_0^1 L(\boldsymbol{\Psi}; {}^{\alpha}\widetilde{\mathbf{x}}) d\alpha = L(\boldsymbol{\Psi}; \widetilde{\mathbf{x}}).$$

In the special case where the complete data $\mathbf{x} = (x_1, \ldots, x_n)$ is a realization of an independent identically distributed (i.i.d.) random vector $\mathbf{X} = (X_1, \ldots, X_n)$, and assuming the joint membership function $\mu_{\widetilde{\mathbf{x}}}$ to be decomposable as in (2), the likelihood function (3) can be written as a product of n terms:

$$L(\boldsymbol{\Psi}; \widetilde{\mathbf{x}}) = \prod_{i=1}^{n} \int \mu_{\widetilde{x}_i}(x) g(x; \boldsymbol{\Psi}) dx, \tag{4}$$

and the observed-data log likelihood is

$$\log L(\boldsymbol{\Psi}; \widetilde{\mathbf{x}}) = \sum_{i=1}^{n} \log \int \mu_{\widetilde{x}_i}(x) g(x; \boldsymbol{\Psi}) dx. \tag{5}$$

[1] We thank an anonymous referee for bringing this point to our attention.

2.3 Fuzzy Information System model

Although the model outlined in Subsection 2.1 will be adopted throughout this paper, it is interesting to compare it with another model that has been extensively used in the fuzzy statistics literature in the 1980's (see, e.g., [16, 14, 15]). In this alternative model, we assume the existence of *fuzzy information system* (FIS) , defined as a fuzzy partition $\mathcal{F} = \{\widetilde{\boldsymbol{\xi}}_1, \ldots, \widetilde{\boldsymbol{\xi}}_K\}$ of $\mathcal{X}$, i.e., a set of K fuzzy subsets of $\mathcal{X}$ verifying the orthogonality condition:

$$\sum_{k=1}^{K} \mu_{\widetilde{\boldsymbol{\xi}}_k}(\mathbf{x}) = 1,$$

where $\mu_{\widetilde{\boldsymbol{\xi}}_k}$ denotes the membership function of $\widetilde{\boldsymbol{\xi}}_k$. When observing a realization $\mathbf{x}$ of $\mathbf{X}$, the experimenter selects the element of $\mathcal{F}$ that best describes his/her perception of $\mathbf{x}$. If we make the (strong) assumption that, conditionally on $\mathbf{x}$, each $\widetilde{\boldsymbol{\xi}}_k$ is selected with probability $\mu_{\widetilde{\boldsymbol{\xi}}_k}(\mathbf{x})$, then the overall probability of selecting $\widetilde{\boldsymbol{\xi}}_k$ is

$$P(\widetilde{\boldsymbol{\xi}}_k) = \int_{\mathcal{X}} \mu_{\widetilde{\boldsymbol{\xi}}_k}(\mathbf{x}) g(\mathbf{x}; \boldsymbol{\Psi}) d\mathbf{x} = \mathbb{E}_{\boldsymbol{\Psi}} \left[\mu_{\widetilde{\boldsymbol{\xi}}_k}(\mathbf{X}) \right], \tag{6}$$

assuming $\mu_{\widetilde{\boldsymbol{\xi}}_k}$ to be Borel measurable. We observe that (6) coincides with Zadeh's definition of the probability of a fuzzy event [35]. However, in the particular model considered here, $P(\widetilde{\boldsymbol{\xi}}_k)$ is the *objective* probability (i.e., the limit frequency) of selecting $\widetilde{\boldsymbol{\xi}}_k$ as the best fuzzy description of $\mathbf{x}$, based on the experimenter's perception. The usual case of categorized data is recovered as a special case when the elements of $\mathcal{F}$ are crisp subsets of $\mathcal{X}$.

Having observed $\widetilde{\mathbf{x}}$, the likelihood function is

$$L(\boldsymbol{\Psi}) = P(\widetilde{\mathbf{X}} = \widetilde{\mathbf{x}}; \boldsymbol{\Psi}) = \int_{\mathcal{X}} \mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) g(\mathbf{x}; \boldsymbol{\Psi}) d\mathbf{x}, \tag{7}$$

which is *formally* identical to (3), with a completely different interpretation. Consequently, the problem of finding the MLE in this model can be addressed in exactly the same way as in the model considered in Subsection 2.1, which will be adopted hereafter for clarity, unless otherwise specified. A solution to this problem is provided in the next section.

3 The Fuzzy EM method

The problem tackled in this section consists in maximizing the observed-data log likelihood, as defined in Section 2.2, in situations where the observed data is fuzzy and can be seen as an incomplete specification of a complete data vector $\mathbf{x}$. The EM algorithm is a broadly applicable mechanism for computing MLEs from incomplete data, in situations where ML estimation would be straightforward if complete data were available [6]. In its basic form, this algorithm assumes that the observed data corresponds to a crisp set of possible values for the complete data. This algorithm thus needs to be adapted to handle fuzzy data. The EM algorithm will first be recalled in Subsection 3.1, and its application to fuzzy data, referred to as the Fuzzy EM (FEM) method, will then be presented in Subsection 3.2.

3.1 The EM algorithm

With the same notations as in Section 2, let us assume that we have a random vector $\mathbf{X}$ with p.d.f. $g(\mathbf{x};\boldsymbol{\Psi})$. A realization $\mathbf{x}$ has been drawn from $\mathbf{X}$, but it is incompletely observed. The observed data consists in a subset $\mathbb{X}$ of $\mathcal{X}$ such that $\mathbf{x} \in \mathbb{X}$. The observed-data likelihood is

$$L(\boldsymbol{\Psi};\mathbb{X}) = \int_{\mathbb{X}} g(\mathbf{x};\boldsymbol{\Psi})d\mathbf{x}. \tag{8}$$

The EM algorithm approaches the problem of maximizing the observed-data log likelihood $\log L(\boldsymbol{\Psi};\mathbb{X})$ by proceeding iteratively with the complete-data log likelihood $\log L(\boldsymbol{\Psi};\mathbf{x}) = \log g(\mathbf{x};\boldsymbol{\Psi})$. Each iteration of the algorithm involves two steps called the expectation step (E-step) and the maximization step (M-step).

The E-step requires the calculation of

$$Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(q)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(q)}}\left[\log L(\boldsymbol{\Psi};\mathbf{X}) \mid \mathbb{X}\right],$$

where $\boldsymbol{\Psi}^{(q)}$ denotes the current fit of $\boldsymbol{\Psi}$ at iteration q, and $\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}$ denotes expectation using the parameter vector $\boldsymbol{\Psi}^{(q)}$.

The M-step then consists in maximizing $Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(q)})$ with respect to $\boldsymbol{\Psi}$ over the parameter space $\boldsymbol{\Omega}$, i.e., finding $\boldsymbol{\Psi}^{(q+1)}$ such that

$$Q(\boldsymbol{\Psi}^{(q+1)},\boldsymbol{\Psi}^{(q)}) \geq Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(q)})$$

for all $\boldsymbol{\Psi} \in \boldsymbol{\Omega}$. The E- and M-steps are iterated until the difference $L(\boldsymbol{\Psi}^{(q+1)};\mathbb{X}) - L(\boldsymbol{\Psi}^{(q)};\mathbb{X})$ becomes smaller than some arbitrarily small amount.

It is proved in [6] that the observed-data likelihood $L(\boldsymbol{\Psi};\mathbb{X})$ is not decreased after an EM iteration, that is,

$$L(\boldsymbol{\Psi}^{(q+1)};\mathbb{X}) \geq L(\boldsymbol{\Psi}^{(q)};\mathbb{X})$$

for $q = 0, 1, 2, \ldots$. Hence, convergence to some value L^* is ensured as long as the sequence $L(\boldsymbol{\Psi}^{(q)};\mathbb{X})$ for $q = 0, 1, 2, \ldots$ is bounded from above. As noted in [26, page 85], L^* is, in most practical applications and except in pathological cases, a local maximum of the incomplete data log likelihood $L(\boldsymbol{\Psi};\mathbb{X})$.

REMARK 1 In [6], Dempster et al. postulate the existence of a sample space $\mathcal{Y}$, and a many-to-one mapping φ from $\mathcal{X}$ to $\mathcal{Y}$. To each observed $\mathbf{y} \in \mathcal{Y}$ thus corresponds a subset $\mathbb{X} = \varphi^{-1}(\mathbf{y})$ of $\mathcal{X}$. If $\mathbf{y}$ is a realization of a random variable $\mathbf{Y}$, then the set $\mathbb{X}$ also becomes random. To be consistent with the model described in Subsection 2.1, the set $\mathbb{X}$ is not considered to be random in this section. This is why $\mathcal{Y}$ and φ are not needed here. The EM algorithm remains unchanged under this interpretation.

3.2 Application to fuzzy data

Let us now assume, as we did in Section 2, that the observed data consist in a fuzzy subset $\widetilde{\mathbf{x}}$ of $\mathcal{X}$, with Borel measurable membership function $\mu_{\widetilde{\mathbf{x}}}$. The observed data likelihood is now given by (3), which is a direct generalization of (8). To maximize

this function, we propose to adapt the EM algorithm as follows. Let the E-step now consist in the calculation of

$$Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(q)}} \left[\log L(\boldsymbol{\Psi}; \mathbf{X}) \mid \widetilde{\mathbf{x}}\right] \tag{9}$$

$$= \frac{\int \mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) \log[L(\boldsymbol{\Psi}; \mathbf{x})] g(\mathbf{x}; \boldsymbol{\Psi}^{(q)}) d\mathbf{x}}{L(\boldsymbol{\Psi}^{(q)}; \widetilde{\mathbf{x}})}, \tag{10}$$

where the expectation of $\log L(\boldsymbol{\Psi}; \mathbf{X})$ is now taken with respect to the conditional p.d.f. of $\mathbf{x}$ given $\widetilde{\mathbf{x}}$, using parameter vector $\boldsymbol{\Psi}^{(q)}$:

$$g(\mathbf{x}|\widetilde{\mathbf{x}}; \boldsymbol{\Psi}^{(q)}) = \frac{\mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) g(\mathbf{x}; \boldsymbol{\Psi}^{(q)})}{\int \mu_{\widetilde{\mathbf{x}}}(\mathbf{u}) g(\mathbf{u}; \boldsymbol{\Psi}^{(q)}) d\mathbf{u}}.$$

(See (20) in Appendix A.1 for the general expression of the p.d.f. of a random variable conditionally on a fuzzy event).

The M-step is unchanged and requires the maximization of $Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)})$ with respect to $\boldsymbol{\Psi}$. The proposed algorithm alternately repeats the E- and M-steps above until the increase of observed-data likelihood becomes smaller than some threshold.

To show that the above algorithm actually maximizes the observed-data likelihood (3), we may remark that it is *formally* equivalent to the standard EM algorithm recalled in Subsection 3.1, applied to a different statistical model with crisp observations, in which the fuzzy data play the role of known parameters[2]. More precisely, let Y denote a Bernoulli random variable such that $P(Y = 1|\mathbf{X} = \mathbf{x}) = \mu_{\widetilde{\mathbf{x}}}(\mathbf{x})$, and let $\mathbf{Z} = (\mathbf{X}, Y)$ be the "complete data". Its density when $Y = 1$ is

$$g(\mathbf{x}, 1; \boldsymbol{\Psi}) = \mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) g(\mathbf{x}; \boldsymbol{\Psi}).$$

Assume that we have observed $Y = 1$. Then the likelihood function based on this observation is

$$L(\boldsymbol{\Psi}; Y = 1) = P(Y = 1; \boldsymbol{\Psi}) = \int g(\mathbf{x}, 1; \boldsymbol{\Psi}) d\mathbf{x} = \int \mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) g(\mathbf{x}; \boldsymbol{\Psi}) d\mathbf{x},$$

which is identical to (3). This likelihood may be maximized using the classical EM algorithm. The E-step of this algorithm computes

$$Q'(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(q)}} \left[\log L(\boldsymbol{\Psi}; \mathbf{Z}) \mid Y = 1\right] \tag{11}$$

$$= \frac{\int \mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) \log[\mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) g(\mathbf{x}; \boldsymbol{\Psi})] g(\mathbf{x}; \boldsymbol{\Psi}^{(q)}) d\mathbf{x}}{\int \mu_{\widetilde{\mathbf{x}}}(\mathbf{u}) g(\mathbf{u}; \boldsymbol{\Psi}^{(q)}) d\mathbf{u}} \tag{12}$$

$$= \frac{\int \mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) \log[\mu_{\widetilde{\mathbf{x}}}(\mathbf{x})] g(\mathbf{x}; \boldsymbol{\Psi}^{(q)}) d\mathbf{x}}{\int \mu_{\widetilde{\mathbf{x}}}(\mathbf{u}) g(\mathbf{u}; \boldsymbol{\Psi}^{(q)}) d\mathbf{u}} + Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)}). \tag{13}$$

As the first term in the right-hand side of (13) does not depend of $\boldsymbol{\Psi}$, it need not be calculated. Maximizing $Q'(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)})$ is equivalent to maximizing $Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)})$, which shows the formal equivalence between the computations under the two models.

The above line of reasoning shows that the Fuzzy EM method introduced in this section can be considered, *from a purely formal point of view*, as the application of the classical EM algorithm to a different statistical model. An immediate consequence is that the proposed procedure generates a nondecreasing sequence of observed-data likelihood values, which converges to some limit if it is bounded.

[2]This argument was suggested to the author by an anonymous referee.

4 Applications

In the previous section, we have shown that the EM algorithm can be adapted to compute MLEs from fuzzy data. In this section, we will now demonstrate the application of this method to three classical estimation problems: mean and variance of univariate normal data (Subsection 4.1), multiple regression with crisp inputs and fuzzy outputs (Subsection 4.2) and univariate normal mixture estimation from fuzzy data (Subsection 4.3).

4.1 Mean and variance of univariate normal data

Let us assume that the complete data $\mathbf{x} = (x_1, \ldots, x_n)'$ is a realization of an i.i.d. random sample from a normal distribution with mean m and standard deviation σ. The observed data is supposed to take the form of a fuzzy subset $\widetilde{\mathbf{x}}$ of $\mathbb{R}^n$ with membership function

$$\mu_{\widetilde{\mathbf{x}}}(\mathbf{x}) = \prod_{i=1}^{n} \mu_{\widetilde{x}_i}(x_i) \tag{14}$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$ and the $\widetilde{x}_i$, $i = 1, \ldots, n$ are fuzzy numbers.

The complete-data p.d.f is

$$g(\mathbf{x}; \boldsymbol{\Psi}) = \prod_{i=1}^{n} g(x_i; \boldsymbol{\Psi}),$$

where $\boldsymbol{\Psi} = (m, \sigma)'$ and

$$g(x_i; \boldsymbol{\Psi}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - m)^2}{2\sigma^2}\right).$$

The complete-data log likelihood is thus

$$\begin{aligned}
\log L(\boldsymbol{\Psi}; \mathbf{x}) &= \sum_{i=1}^{n} \log g(x_i; \boldsymbol{\Psi}) \\
&= -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2 \\
&= -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2m\sum_{i=1}^{n} x_i + nm^2\right).
\end{aligned}$$

The observed-data log likelihood is given by (5).

At iteration $q+1$, the E-step of the EM algorithm requires the calculation of

$$\begin{aligned}
Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(q)}}\left[\log L(\boldsymbol{\Psi}; \mathbf{X}) | \widetilde{\mathbf{x}}\right] &= -\frac{n}{2}\log(2\pi) - n\log\sigma \\
&- \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(X_i^2 | \widetilde{x}_i) - 2m\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(X_i | \widetilde{x}_i) + nm^2\right).
\end{aligned}$$

When the $\widetilde{x}_i$ are trapezoidal fuzzy numbers, the conditional expectations $\alpha_i^{(q)} = \mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_i^2|\widetilde{x}_i)$ and $\beta_i^{(q)} = \mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_i|\widetilde{x}_i)$ can be computed using Equation (23)-(28) in Appendix A.2.

The M-step of the EM algorithm involves maximizing $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})$ with respect to $\mathbf{\Psi}$. This is easily achieved by solving the likelihood equations. From

$$\frac{\partial Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})}{\partial m} = -2\sum_{i=1}^{n} \beta_i^{(q)} + 2nm,$$

we get

$$m^{(q+1)} = \frac{1}{n}\sum_{i=1}^{n} \beta_i^{(q)}.$$

Now,

$$\frac{\partial Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left(\sum_{i=1}^{n} \alpha_i^{(q)} - 2m\sum_{i=1}^{n} \beta_i^{(q)} + nm^2\right).$$

Substituting $m^{(q+1)}$ for m in the equation

$$\frac{\partial Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})}{\partial \sigma} = 0$$

and solving for σ, we get

$$\sigma^{(q+1)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \alpha_i^{(q)} - (m^{(q+1)})^2}.$$

The EM algorithm for this problem is summarized in Algorithm 1.

EXAMPLE 2 To illustrate the above algorithm, we consider real data[3] collected during the experiment reported in [20]. In this experiment, subjects were asked about their perception of the relative length of different line segments with respect to a fixed longer segment that was used as a standard for comparison. The subjects had to describe their perception by providing the support and the core of trapezoidal fuzzy numbers. In the version of the database used in this example, there were 17 subjects, and segments with various relative lengths were presented in random order, with three repetitions for each length. We considered the $n = 17 \times 3 = 51$ fuzzy numbers corresponding to the true relative length 61.47. Figure 1 shows the first 10 of these trapezoidal fuzzy numbers.

The model described in Subsection 2.1 can be applied to this data if we assume that each reported fuzzy number $\widetilde{x}_i$ acts a soft constraint on the observer's perceived relative length x_i, considered to be a realization of a random variable $X_i = m + \varepsilon$, where $m = 61.47$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Note that this model is different from that used in [20], in which $\widetilde{x}_i$ is treated as a realization of a fuzzy random variable. A detailed comparison between these two models would go beyond the scope of this paper.

[3]The dataset can be downloaded from the webpage of the SMIRE research group at http://bellman.ciencias.uniovi.es/SMIRE/Perceptionsdata.html.

Algorithm 1 EM algorithm for the estimation of the mean and standard deviation of a normal population using a fuzzy random sample.

Input: $\widetilde{x}_1, \ldots, \widetilde{x}_n$, $\epsilon > 0$.
Output: $\widehat{m}$, $\widehat{\sigma}$

$q \leftarrow 0$
Initialization: pick $m^{(0)}$ and $\sigma^{(0)}$
$L(\mathbf{\Psi}^{(0)}; \widetilde{\mathbf{x}}) \leftarrow \sum_{i=1}^{n} \log \mathbb{E}_{\mathbf{\Psi}^{(0)}}(\mu_{\widetilde{x}_i}(X_i))$ *% using (21)-(22)*
repeat
  % E-step
  for $i = 1 : n$ **do**
    $\alpha_i^{(q)} \leftarrow \mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_i^2 | \widetilde{x}_i)$ *% using (21)-(28)*
    $\beta_i^{(q)} \leftarrow \mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_i | \widetilde{x}_i)$ *% using (21)-(25)*
  end for
  % M-step
  $m^{(q+1)} \leftarrow \frac{1}{n} \sum_{i=1}^{n} \beta_i^{(q)}$
  $\sigma^{(q+1)} \leftarrow \sqrt{\frac{1}{n} \sum_{i=1}^{n} \alpha_i^{(q)} - (m^{(q+1)})^2}$
  $L(\mathbf{\Psi}^{(q+1)}; \widetilde{\mathbf{x}}) \leftarrow \sum_{i=1}^{n} \log \mathbb{E}_{\mathbf{\Psi}^{(q+1)}}(\mu_{\widetilde{x}_i}(X_i))$ *% using (21)-(22)*
  $D \leftarrow (L(\mathbf{\Psi}^{(q+1)}; \widetilde{\mathbf{x}}) - L(\mathbf{\Psi}^{(q)}; \widetilde{\mathbf{x}}))/|L(\mathbf{\Psi}^{(q)}; \widetilde{\mathbf{x}})|$
  $q \leftarrow q + 1$
until $D < \epsilon$
$\widehat{m} \leftarrow m^{(q)}$; $\widehat{\sigma} \leftarrow \sigma^{(q)}$

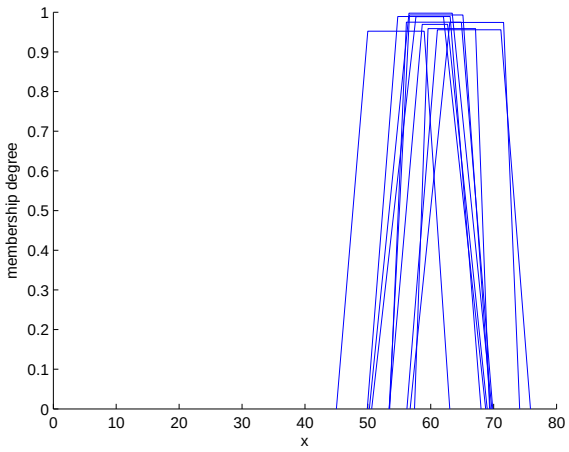

Figure 1: Ten trapezoidal fuzzy numbers considered in Example 2. The heights have been jittered to better separate the membership functions.

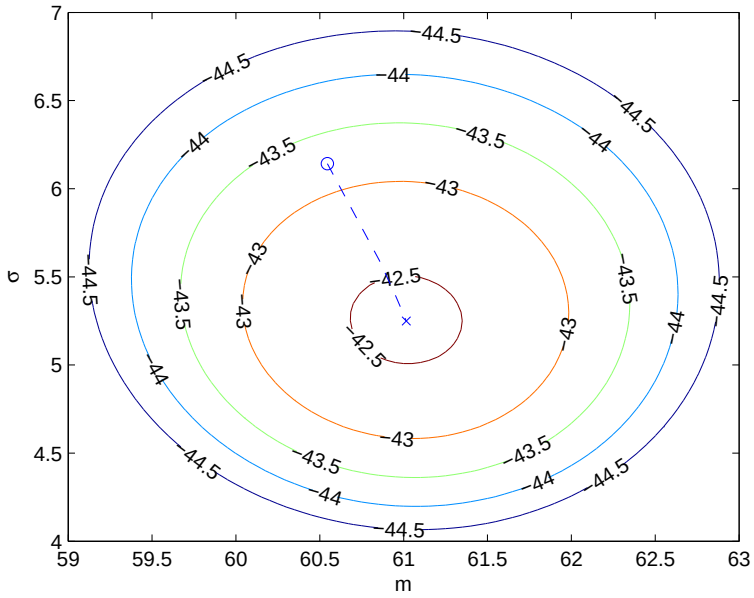

Figure 2: Contour plot of the observed data log likelihood function and trajectory in parameter space (broken line) from the initial parameter values $(m^{(0)}, \sigma^{(0)})$ (o) to the final MLE $(\widehat{m}, \widehat{\sigma})$ (x), for the data of Example 2.

As initial estimates of m and σ, we used the sample mean and standard deviations computed over the centers of cores of each trapezoidal fuzzy numbers, equal to $m^{(0)} = 60.5455$ and $\sigma^{(0)} = 6.1423$. The EM converged after 10 iterations. The stopping criterion was based on the relative change of the log likelihood, with a tolerance value $\epsilon = 10^{-8}$. The final MLEs were $\widehat{m} = 61.0155$ and $\widehat{\sigma} = 5.2494$.

Figure 2 shows a contour plot of the observed data log likelihood function as well as the trajectory in parameter space from the initial parameter values $(m^{(0)}, \sigma^{(0)})$ to the final MLE $(\widehat{m}, \widehat{\sigma})$. We can check that the MLE corresponds in this case to a global maximum of the observed data log likelihood. In more complex problem, the algorithm may be trapped in local maxima. It is then necessary to start it several times with different random initial conditions. □

Example 3 To illustrate experimentally the asymptotic behavior of the MLEs of m and σ under the model described in Subsection 2.3 (which lends itself easily to random simulation), we performed the following experiment. We considered i.i.d random samples of size n from the standard normal distribution. Each realization of $\mathbf{x}$ was fuzzified using the FIS shown in Figure 3, and the MLEs $\widehat{m}$ and $\widehat{\sigma}$ for the fuzzy sample were computed using the FEM method. For each value of n (ranging from 10 to 1000), the whole procedure was repeated 100 times and the expectation and standard error of $\widehat{m}$ and $\widehat{\sigma}$ were estimated. The results are shown in Figures 4 and 5, which illustrate the convergence of both estimators towards the true parameter values. We note that similar results are obtained if each fuzzy number $\widetilde{\xi}_k$ in the FIS is replaced by its cut at level 0.5, defining a crisp partition of the real line. □

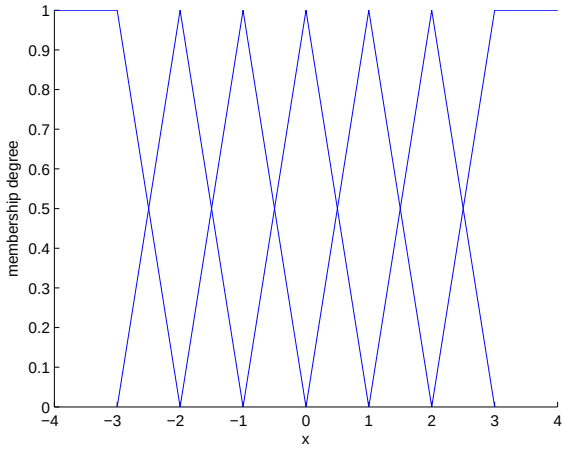

Figure 3: Fuzzy information system used to encode the simulated fuzzy data of Example 3.

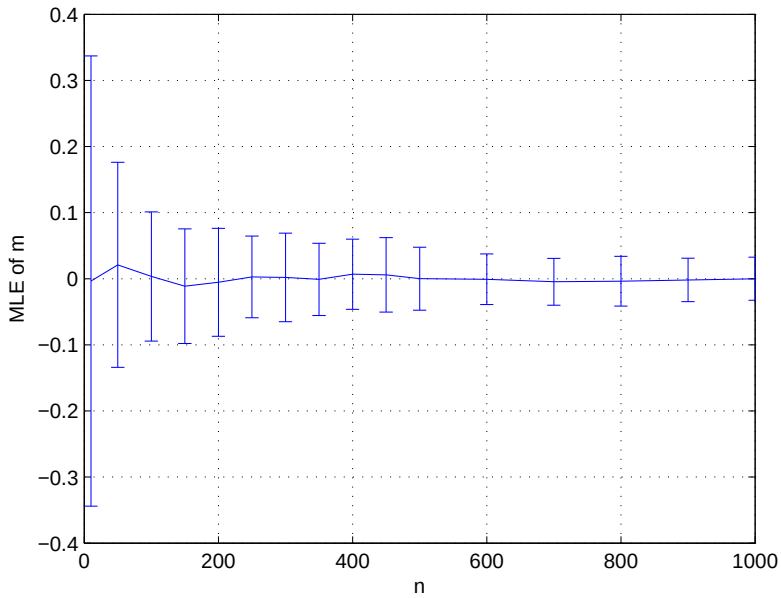

Figure 4: Mean plus or minus one standard deviation (over 100 trials) of $\widehat{m}$ for a fuzzy sample of size n from a standard normal distribution, as a function of n (Example 3).

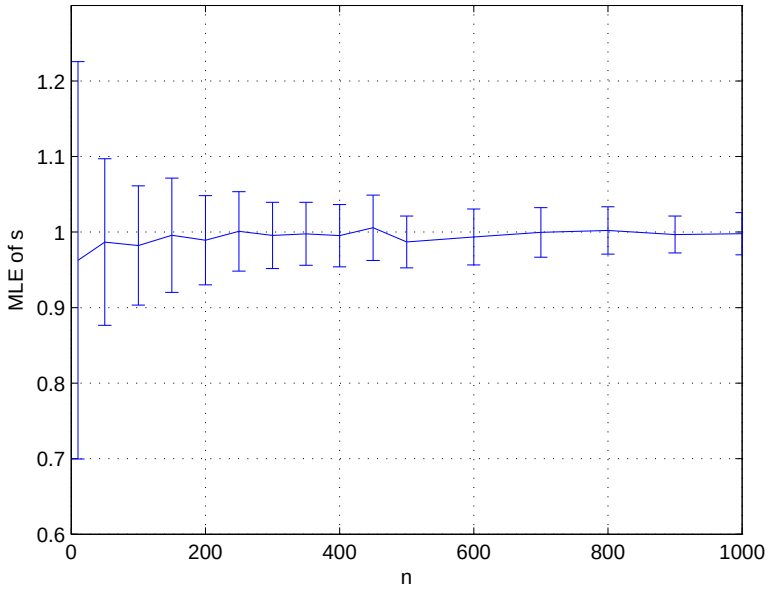

Figure 5: Mean plus or minus one standard deviation (over 100 trials) of $\widehat{\sigma}$ for a fuzzy sample of size n from a standard normal distribution, as a function of n (Example 3).

4.2 Multiple regression

Let us now consider the case of multiple regression with crisp inputs and fuzzy outputs. This is an old problem in fuzzy data analysis (see, e.g., [4, 27] and references therein). As we will show in this section, the FEM method provides a new and well motivated solution to this problem, with very simple implementation.

The fuzzy data $\widetilde{\mathbf{x}}$ will be assumed to have the same form as in the previous section. However, we now assume that each component x_i of the complete data vector $\mathbf{x}$ is a realization of a normal random variable X_i with mean $\mathbf{u}_i'\mathbf{b}$ and standard deviation σ, where $\mathbf{u}_i = (1, u_{i1}, \ldots, u_{i,p-1})'$ is a constant p-dimensional input vector and $\mathbf{b}$ is vector of p coefficients[4]. The complete parameter vector is thus $\mathbf{\Psi} = (\mathbf{b}, \sigma)'$.

Denoting by $\mathbf{U}$ the matrix of n rows and p columns, with row i equal to $\mathbf{u}_i$, and assuming independence between the X_i, the observed data vector $\mathbf{x}$ is multivariate Gaussian with mean $\mathbf{Ub}$ and variance $\sigma^2 I_p$, where I_p denotes the p-dimensional identity matrix. The complete-data log likelihood is thus:

$$\begin{aligned} \log L(\mathbf{\Psi}; \mathbf{x}) &= -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}(\mathbf{x}-\mathbf{Ub})'(\mathbf{x}-\mathbf{Ub}) \\ &= -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\left(\mathbf{x}'\mathbf{x} - 2\mathbf{b}'\mathbf{U}'\mathbf{x} + \mathbf{b}'\mathbf{U}'\mathbf{Ub}\right). \end{aligned}$$

Taking the expectation of $\log L(\mathbf{\Psi}; \mathbf{X})$ conditionally on the observed fuzzy data and

[4]To be consistent with our previous notations, we have to depart from the usual convention in regression analysis, where the dependent and independent variables are denoted by Y and $\mathbf{x}$, respectively. We hope that the reader will not be confused by this change of notation.

using the fit $\mathbf{\Psi}^{(q)}$ of $\mathbf{\Psi}$ to perform the E-step, we get

$$Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)}) = \mathbb{E}_{\mathbf{\Psi}^{(q)}}\left[\log L(\mathbf{\Psi}; \mathbf{X})|\widetilde{\mathbf{x}}\right] = -\frac{n}{2}\log(2\pi) - n\log\sigma \\ - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_i^2|\widetilde{x}_i) - 2\mathbf{b}'\mathbf{U}'\mathbb{E}_{\mathbf{\Psi}^{(q)}}(\mathbf{X}|\widetilde{\mathbf{x}}) + \mathbf{b}'\mathbf{U}'\mathbf{U}\mathbf{b}\right). \quad (15)$$

As before, let us denote $\alpha_i^{(q)} = \mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_i^2|\widetilde{x}_i)$, $\beta_i^{(q)} = \mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_i|\widetilde{x}_i)$ and

$$\boldsymbol{\beta}^{(q)} = \mathbb{E}_{\mathbf{\Psi}^{(q)}}(\mathbf{X}|\widetilde{\mathbf{x}}) = (\mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_1|\widetilde{x}_1), \ldots, \mathbb{E}_{\mathbf{\Psi}^{(q)}}(X_n|\widetilde{x}_n))'.$$

With these notations, (15) becomes:

$$Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)}) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\alpha_i^{(q)} - 2\mathbf{b}'\mathbf{U}'\boldsymbol{\beta}^{(q)} + \mathbf{b}'\mathbf{U}'\mathbf{U}\mathbf{b}\right).$$

The M-step requires maximizing $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})$ with respect to $\mathbf{\Psi}$. This can be achieved by differentiating $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})$ with respect to $\mathbf{b}$ and σ, which results in

$$\frac{\partial Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})}{\partial \mathbf{b}} = -\frac{1}{\sigma^2}\left(-\mathbf{U}'\boldsymbol{\beta}^{(q)} + \mathbf{U}'\mathbf{U}\mathbf{b}\right)$$

and

$$\frac{\partial Q(\mathbf{\Psi}, \mathbf{\Psi}^{(q)})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left(\sum_{i=1}^{n}\alpha_i^{(q)} - 2\mathbf{b}'\mathbf{U}'\boldsymbol{\beta}^{(q)} + \mathbf{b}'\mathbf{U}'\mathbf{U}\mathbf{b}\right).$$

Equating these derivatives to zero and solving for $\mathbf{b}$ and σ, we get the following unique solution:

$$\mathbf{b}^{(q+1)} = (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\boldsymbol{\beta}^{(q)}$$

and

$$\begin{aligned}\sigma^{(q+1)} &= \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}\alpha_i^{(q)} - 2\,\mathbf{b}^{(q+1)\prime}\mathbf{U}'\boldsymbol{\beta}^{(q)} + \mathbf{b}^{(q+1)\prime}\mathbf{U}'\mathbf{U}\mathbf{b}^{(q+1)}\right)} \\ &= \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}\alpha_i^{(q)} - 2\,\widehat{\mathbf{x}}^{(q+1)\prime}\boldsymbol{\beta}^{(q)} + \widehat{\mathbf{x}}^{(q+1)\prime}\widehat{\mathbf{x}}^{(q+1)}\right)},\end{aligned}$$

with $\widehat{\mathbf{x}}^{(q+1)} = \mathbf{U}\mathbf{b}^{(q+1)}$.

Example 4 The above algorithm was applied to the perception data reported in [20] and already described in Example 2. We considered the responses of subject 12 for the 9 different relative lengths. There were 3 repetitions for each relative length, yielding $n = 9 \times 3 = 27$ observations. Here the true relative length and the perceived length are taken as the independent and dependent variables, respectively. The data are shown in Figure 6. The EM algorithm was initialized as in Example 2, with the same stopping criterion. The initial values were $\mathbf{b}^{(0)} = (-5.0942, 0.9919)'$ and $\sigma^{(0)} = 6.6158$. The algorithm converged to the MLE $\widehat{\mathbf{b}} = (-3.9832, 0.9916)'$ and $\widehat{\sigma} = 4.9218$ in 13 iterations.

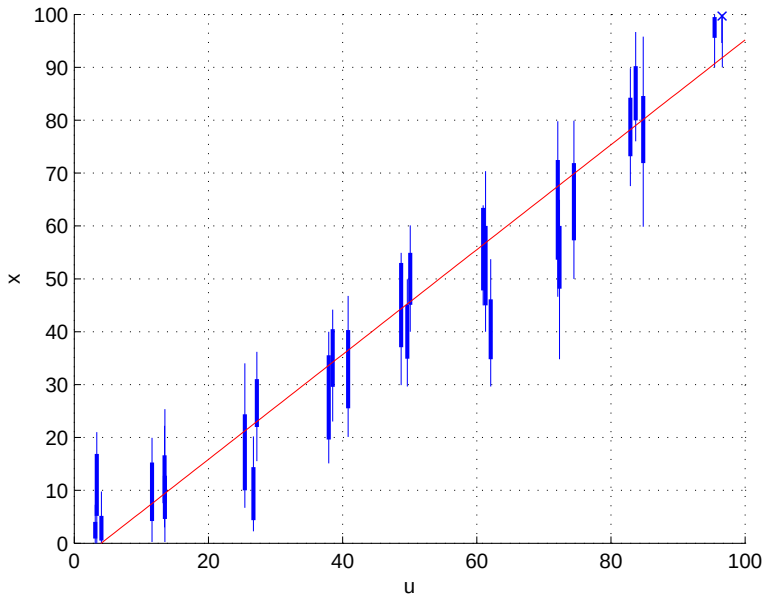

Figure 6: Data of Example 4, and fit of the regression line. The support and core of each fuzzy number are represented by solid and bold line segments, respectively. Jitter has been added to the u_i to avoid superposition of fuzzy numbers.

EXAMPLE 5 To demonstrate the interest of expressing partial information about ill-known data in the form of possibility distributions, as explained in Subsection 2.1, we performed the following experiment. We generated $n = 100$ values $\mathbf{u}_i$ from the uniform distribution in $[0, 2]$, and we generated corresponding values x_i using the linear regression model with $\mathbf{b} = (2, 1)'$ and $\sigma = 0.2$. To model the situation where only partial knowledge of values $x_1, \ldots, x_n$ is available, triangular fuzzy numbers $\widetilde{x}_1, \ldots, \widetilde{x}_n$ were generated as follows:

- For each i, a "guess" x'_i was randomly generated from a normal distribution with mean x_i and standard deviation σ_i, were σ_i was drawn randomly from a uniform distribution in $[0, 0.5]$;
- $\widetilde{x}_i$ was defined as the triangular fuzzy number with core x'_i and support $[x'_i - 2\sigma_i, x'_i + 2\sigma_i]$.

This procedure simulates the situation where the observer has only approximate knowledge of the data, and can only provide a guess x'_i and an interval of plausible values $[x'_i - 2\sigma_i, x'_i + 2\sigma_i]$. It should be emphasized that, although the $\widetilde{x}_i$ are randomly generated for the purpose of this simulation experiment, the relationship between x_i and $\widetilde{x}_i$ is not considered to be random under the model considered here (as explained in Subsection 2.1).

Three strategies were compared for estimating the parameter vector $\mathbf{\Psi} = (\mathbf{b}, \sigma)'$:

1. Using the fuzzy data $\widetilde{x}_1, \ldots, \widetilde{x}_n$ (method 1).

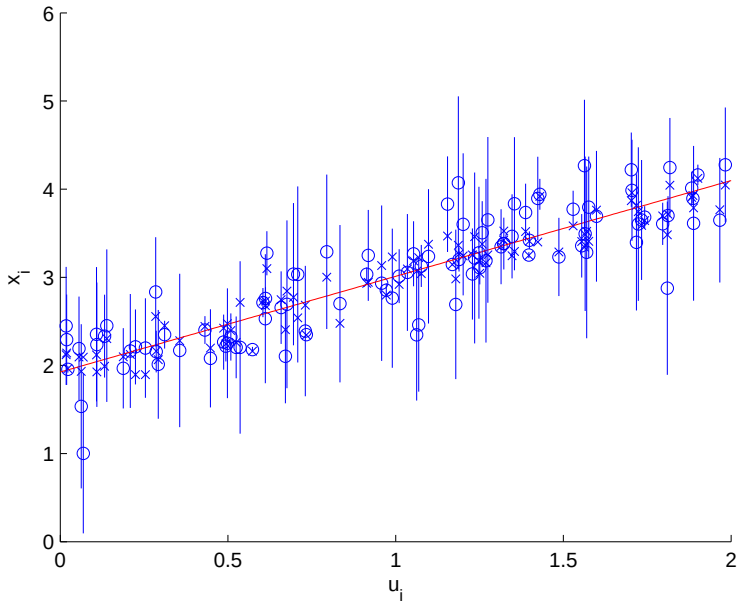

Figure 7: Example of randomly generated data for the experiment of Example 5, with the fit of the regression line. The points (u_i, x_i) and (u_i, x_i') are represented by crosses and circles, respectively. The vertical segments represent the supports of the triangular fuzzy numbers $\widetilde{x}_i$.

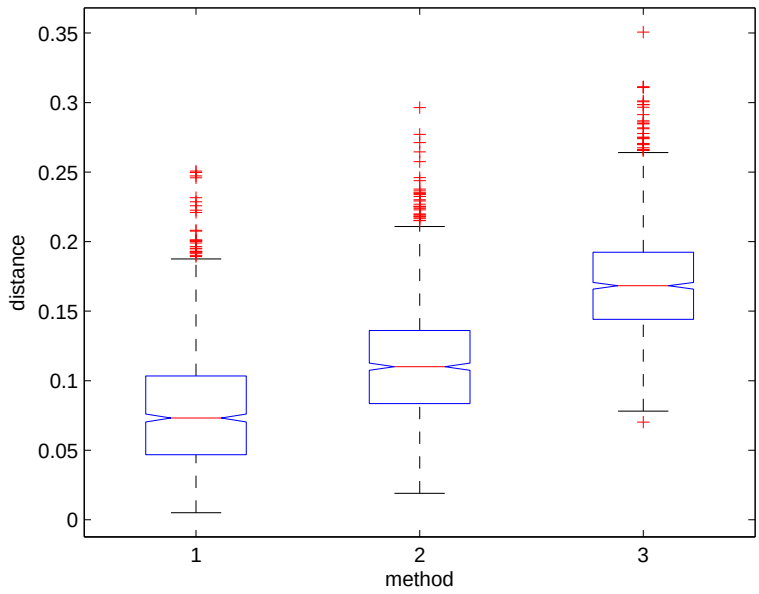

Figure 8: Box plots of the distributions of $\|\widehat{\mathbf{\Psi}} - \mathbf{\Psi}\|$ over 1000 trials for the three methods and the experiment described in Example 5. Boxes whose notches do not overlap indicate that the medians of the two groups differ at the 5% significance level.

2. Using only 0.5-cuts of the fuzzy data, i.e., the interval data ${}^{0.5}\widetilde{x}_1, \ldots, {}^{0.5}\widetilde{x}_n$ (method 2);

3. Using only the crisp guesses $x'_1, \ldots, x'_n$(method 3).

For each of these three methods, the L_2 distance $\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\|$ between the true parameter vector and its MLE was computed. The whole experiment was repeated 1000 times. Results are shown as box plots in Figure 8. As we can see, method 1 using fuzzy assessments of ill-known data and the FEM method indeed yields better estimates than those obtained using interval or crisp data. □

4.3 Univariate normal mixture with common unknown variance

As a third and last illustration of the FEM method, we will consider a more complex situation where a fuzzy sample is seen as an imprecise specification of a crisp sample from a univariate normal mixture with unknown means, variance and mixing proportions. A similar problem was addressed in [22] using the standard EM algorithm, with interval data in place of fuzzy data.

As before, the observed data vector will be assumed to consist of n fuzzy numbers, denoted here $\widetilde{w}_1, \ldots, \widetilde{w}_n$. Each fuzzy number $\widetilde{w}_i$ is interpreted as a possibility distribution constraining an unknown value w_i. The joint possibility distribution of the n values $w_1, \ldots, w_n$ is identified with the fuzzy subset $\widetilde{\mathbf{w}}$ of $\mathbb{R}^n$ with membership function

$$\mu_{\widetilde{\mathbf{w}}}(\mathbf{w}) = \prod_{i=1}^{n} \mu_{\widetilde{w}_i}(w_i),$$

for all $\mathbf{w} = (w_1, \ldots, w_n)$. The n values $w_1, \ldots, w_n$ are assumed to be a realization of an i.i.d. random sample $W_1, \ldots, W_n$ from a finite normal mixture with p.d.f.

$$g(w; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_k g_k(w; \boldsymbol{\theta}_k),$$

where $g_k(w; \boldsymbol{\theta}_k)$ is a normal p.d.f. with parameters $\boldsymbol{\theta}_k = (m_k, \sigma_k)'$, π_k is the mixing proportion of the k-th component, g is the number of components, and

$$\boldsymbol{\Psi} = (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_g, \pi_1, \ldots, \pi_g)'$$

is the vector of parameters.

Using a classical device when handling finite mixture problems using the EM algorithm, let us introduce a vector $\mathbf{z} = (\mathbf{z}'_1, \ldots, \mathbf{z}'_n)'$, where $\mathbf{z}_i$ is a vector or zero-one indicator variables such that $z_{ik} = 1$ if w_i arose from the k-th component of the mixture, and $z_{ik} = 0$ otherwise. The complete data vector is thus

$$\mathbf{x} = (\mathbf{w}', \mathbf{z}')',$$

and the complete-data p.d.f. is

$$g(\mathbf{x}; \boldsymbol{\Psi}) = \prod_{i=1}^{n} g(\mathbf{z}_i; \boldsymbol{\Psi}) g(\mathbf{w}|\mathbf{z}_i; \boldsymbol{\Psi}) = \prod_{i=1}^{n} \left(\prod_{k=1}^{g} {\pi_k}^{z_{ik}} \right) \left(\prod_{k=1}^{g} g_k(w; \boldsymbol{\theta}_k)^{z_{ik}} \right),$$

from which we can deduce the expression of the complete-data log likelihood:

$$\begin{aligned}\log L(\boldsymbol{\Psi};\mathbf{x}) &= \sum_{i=1}^{n}\sum_{k=1}^{g} z_{ik}\log\pi_k + \sum_{i=1}^{n}\sum_{k=1}^{g} z_{ik}\log g_k(w_i;\boldsymbol{\theta}_k)\\ &= \sum_{k=1}^{g}\log\pi_k\sum_{i=1}^{n} z_{ik} - \frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{k=1}^{g} z_{ik}(w_i-m_k)^2.\end{aligned}$$

To perform the E-step of the EM algorithm, we need to compute the conditional expectation of the complete-data log likelihood conditionally on the observed data $\widetilde{\mathbf{w}}$, using the current fit $\boldsymbol{\Psi}^{(q)}$ of the parameter vector:

$$\begin{aligned}Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(q)}) &= \mathbb{E}_{\boldsymbol{\Psi}^{(q)}}\left[\log L(\boldsymbol{\Psi};\mathbf{X}) \mid \widetilde{\mathbf{w}}\right]\\ &= \sum_{k=1}^{g}\log\pi_k\sum_{i=1}^{n}\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}|\widetilde{w}_i) - \frac{n}{2}\log(2\pi) - n\log\sigma\\ &- \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{k=1}^{g}\left(\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}W_i^2|\widetilde{w}_i) - 2m_k\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}W_i|\widetilde{w}_i) + m_k^2\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}|\widetilde{w}_i)\right). \quad (16)\end{aligned}$$

We thus have to compute the three conditional expectations $\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}|\widetilde{w}_i)$, $\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}W_i^2|\widetilde{w}_i)$ and $\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}W_i|\widetilde{w}_i)$. Let $t_{ik}^{(q)} = \mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}|\widetilde{w}_i)$. From Bayes' theorem, we have

$$t_{ik}^{(q)} = P_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}=1|\widetilde{w}_i) = \frac{P_{\boldsymbol{\Psi}^{(q)}}(\widetilde{w}_i|Z_{ik}=1)P_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}=1)}{P_{\boldsymbol{\Psi}^{(q)}}(\widetilde{w}_i)} = \frac{\gamma_{ik}^{(q)}\pi_k^{(q)}}{p_i^{(q)}},$$

with

$$\gamma_{ik}^{(q)} = \int \mu_{\widetilde{w}_i}(w)g_k(w;\boldsymbol{\theta}_k^{(q)})dw$$

and

$$p_i^{(q)} = \sum_{k=1}^{g}\pi_k^{(q)}\int \mu_{\widetilde{w}_i}(w)g_k(w;\boldsymbol{\theta}_k^{(q)})dx = \sum_{k=1}^{g}\pi_k^{(q)}\gamma_{ik}^{(q)}.$$

Now,

$$\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}W_i^2|\widetilde{w}_i) = \mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(W_i^2|\widetilde{w}_i, Z_{ik}=1)P_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}=1|\widetilde{w}_i) = \xi_{ik}^{(q)}t_{ik}^{(q)},$$

with

$$\xi_{ik}^{(q)} = \mathbb{E}_{\boldsymbol{\theta}_k^{(q)}}(W_i^2|z_{ik}=1,\widetilde{w}_i) = \frac{1}{\gamma_{ik}^{(q)}}\int w^2\mu_{\widetilde{w}_i}(w)g_k(w;\boldsymbol{\theta}_k^{(q)})dw.$$

Similarly,

$$\mathbb{E}_{\boldsymbol{\Psi}^{(q)}}(Z_{ik}W_i|\widetilde{w}_i) = \eta_{ik}^{(q)}t_{ik}^{(q)}$$

with

$$\eta_{ik}^{(q)} = \mathbb{E}_{\boldsymbol{\theta}_k^{(q)}}(W_i|Z_{ik}=1,\widetilde{w}_i) = \frac{1}{\gamma_{ik}^{(q)}}\int w\mu_{\widetilde{w}_i}(w)g_k(w;\boldsymbol{\theta}_k^{(q)})dw.$$

The expected complete-data log-likelihood can thus be written as:

$$Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)}) = \sum_{k=1}^{g} \log \pi_k \sum_{i=1}^{n} t_{ik}^{(q)} - \frac{n}{2}\log(2\pi) - n\log\sigma \\ - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{k=1}^{g}\left[t_{ik}^{(q)}\left(\xi_{ik}^{(q)} - 2m_k\eta_{ik}^{(q)} + m_k^2\right)\right]. \quad (17)$$

Let us now consider the M-step. Maximizing the first term on the right-hand side of (15) with respect to the mixing proportions π_k, under the constraint $\sum_k \pi_k = 1$ yields the classical solution

$$\pi_k^{(q+1)} = \frac{\sum_{i=1}^{n} t_{ik}^{(q)}}{\sum_{\ell=1}^{g}\sum_{i=1}^{n} t_{i\ell}^{(q)}} = \frac{1}{n}\sum_{i=1}^{n} t_{ik}^{(q)}.$$

Now,

$$\frac{\partial Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)})}{\partial m_k} = -\frac{1}{\sigma^2}\left(-\sum_{i=1}^{n} t_{ik}^{(q)}\eta_{ik}^{(q)} + m_k\sum_{i=1}^{n} t_{ik}^{(q)}\right).$$

Equating this derivative to zero and solving for m_k yields:

$$m_k^{(q+1)} = \frac{\sum_{i=1}^{n} t_{ik}^{(q)}\eta_{ik}^{(q)}}{\sum_{i=1}^{n} t_{ik}^{(q)}}.$$

Finally, the derivative of $Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)})$ with respect to σ is

$$\frac{\partial Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(q)})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\sum_{k=1}^{g}\left[t_{ik}^{(q)}\left(\xi_{ik}^{(q)} - 2m_k\eta_{ik}^{(q)} + m_k^2\right)\right].$$

Equating this derivative to zero and substituting $m_k^{(q+1)}$ for m_k, we get the following fit for σ at iteration $q+1$:

$$\sigma^{(q+1)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{g}\left[t_{ik}^{(q)}\left(\xi_{ik}^{(q)} - 2m_k^{(q+1)}\eta_{ik}^{(q)} + (m_k^{(q+1)})^2\right)\right]}.$$

EXAMPLE 6 To illustrate the behavior of the above algorithm, we used again the perception dataset reported in [20] and already described in Examples 2 and 4. We considered the assessments of 8 subjects for the relative lengths around 61.47, 50 and 38.3 (the relative lengths may vary slightly because of screen resolution, as explained in [20]). The considered dataset was thus composed of $n = 72$ trapezoidal fuzzy sets, partitioned in three classes of equal size. A subset of the data is shown in Figure 9. The above algorithm was applied to this dataset with $g = 3$. The means m_k were initialized randomly using a uniform distribution in [0,100], while the initial standard deviation and proportions were set to fixed values as shown in Table 1. The EM algorithm was run 10 times, and the results corresponding to the best value of the objective criterion were retained. The MLE estimates are shown in Table 1, and the corresponding mixture density estimated is shown in Figure 9. We can see that the three relative length estimates are very close to their true values.

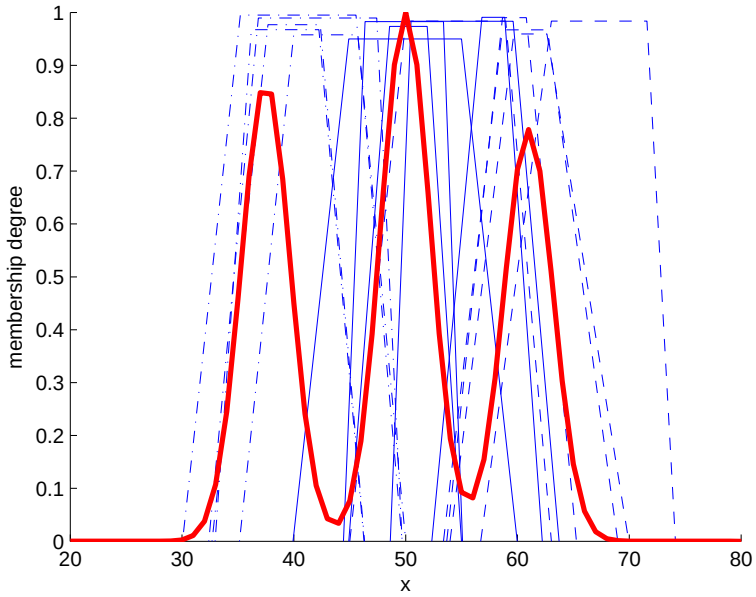

Figure 9: Subset of the data of Example 6 (with 5 fuzzy numbers in each class), and normalized estimated mixture p.d.f. (bold line). The heights of fuzzy membership functions have been jittered for better visualization.

Table 1: Initial parameter values and MLEs with $g = 3$ components for the data of Example 6.

parameter	m_1	m_2	m_3	σ	π_1	π_2	π_3
initial value	89.2054	12.7208	78.3867	10	0.3333	0.3333	0.3333
MLE	60.9896	37.4851	49.9999	2.1915	0.2940	0.3283	0.3777

To conclude this section, we can remark that determining the number of components in a mixture model is an important problem that has been addressed by many authors (see, e.g., [31, 1]). Some approaches are based on a Bayesian approach, which results in the addition of a regularization term to the log-likelihood criterion. Whether these or other approaches can be extended to the context of fuzzy data is an interesting question that goes beyond the scope of this paper and is left for further study.

5 Conclusions

We have shown that the EM algorithm can be adapted to handle estimation problems in which the observed data are fuzzy and are assumed to be related to underlying crisp data, making it possible to implement the maximum likelihood principle in this context. The proposed FEM method is very general and can be applied to a wide range of statistical problems. In this paper, it has been applied to classical parameter estimation tasks including multiple regression analysis with crisp input data and fuzzy output data, and univariate finite normal mixture estimation from fuzzy data. More complex problems such as principal component analysis [18, 9], discriminant analysis, and clustering [3, 12, 30] of multidimensional fuzzy data can be handled by this method as well. Another direction of research concerns the application of a similar methodology to more general types of imprecise and uncertain data such as described, e.g., by belief functions. Preliminary results in this direction have been reported in [7].

Acknowledgement

The author thanks the members of SMIRE group and, in particular, Ana Colubi, María Ángeles Gil and Gil González-Rodríguez, for making the perception dataset publicly available. He also expresses his thanks to the two anonymous referees for their constructive and very helpful comments.

A Probability of fuzzy events

The notion of probability was extended to fuzzy events by Zadeh [35]. In this appendix, we recall the main definitions and we derive some of the formula used in the paper.

A.1 Basic definitions

Let $(\mathbb{R}^n, \mathcal{A}, P)$ be a probability space in which $\mathcal{A}$ is the σ-field of Borel sets and P is a probability measure on $\mathbb{R}^n$. Then a fuzzy event in $\mathbb{R}^n$ is a fuzzy subset $\widetilde{A}$ of $\mathbb{R}^n$ whose membership function $\mu_{\widetilde{A}}$ is Borel measurable. The probability of $\widetilde{A}$ is defined as the expectation of $\mu_{\widetilde{A}}$ with respect to P:

$$P(\widetilde{A}) = \int \mu_{\widetilde{A}}(\mathbf{x}) dP. \tag{18}$$

Two fuzzy events $\widetilde{A}$ and $\widetilde{B}$ in the probability space $(\mathbb{R}^n, \mathcal{A}, P)$ are said to be independent if

$$P(\widetilde{A}\widetilde{B}) = P(\widetilde{A})P(\widetilde{B}),$$

where $\widetilde{A}\widetilde{B}$ is the fuzzy subset of $\mathbb{R}^n$ with membership function

$$\mu_{\widetilde{A}\widetilde{B}}(\mathbf{x}) = \mu_{\widetilde{A}}(\mathbf{x}) \cdot \mu_{\widetilde{B}}(\mathbf{x}),$$

for all x in $\mathbb{R}^n$. The conditional probability of $\widetilde{A}$ given $\widetilde{B}$ is defined by

$$P(\widetilde{A}|\widetilde{B}) = \frac{P(\widetilde{A}\widetilde{B})}{P(\widetilde{B})}, \tag{19}$$

provided $P(\widetilde{B}) > 0$).

In particular, assume that P is the probability distribution of a continuous random variable $\mathbf{X}$ with p.d.f. $g(\mathbf{x})$. For a crisp subset A and a fuzzy subset $\widetilde{B}$, Equation (19) becomes

$$P(A|\widetilde{B}) = \frac{\int \mu_A(\mathbf{x})\mu_{\widetilde{B}}(\mathbf{x})g(\mathbf{x})d\mathbf{x}}{\int \mu_{\widetilde{B}}(\mathbf{u})g(\mathbf{u})d\mathbf{u}} = \int_A \frac{\mu_{\widetilde{B}}(\mathbf{x})g(\mathbf{x})}{\int \mu_{\widetilde{B}}(\mathbf{u})g(\mathbf{u})d\mathbf{u}}d\mathbf{x}.$$

The conditional density of $\mathbf{X}$ given $\widetilde{B}$ can thus be defined as:

$$g(\mathbf{x}|\widetilde{B}) = \frac{\mu_{\widetilde{B}}(\mathbf{x})g(\mathbf{x})}{\int \mu_{\widetilde{B}}(\mathbf{u})g(\mathbf{u})d\mathbf{u}}. \tag{20}$$

A.2 Trapezoidal fuzzy events and univariate normal distribution

To illustrate the above definitions, let us assume that P is the distribution of a univariate normal random variable X with mean m, standard deviation σ and p.d.f. $g(x)$. Let $\widetilde{x} = (a, b, c, d)$ be a trapezoidal fuzzy number, with membership function

$$\mu_{\widetilde{x}}(x) = \begin{cases} \frac{x-a}{b-a} & \text{if } a \le x \le b, \\ 1 & \text{if } b \le x \le c, \\ \frac{d-x}{d-c} & \text{if } c \le x \le d, \\ 0 & \text{otherwise.} \end{cases}$$

Denoting by $g(x)$ the p.d.f. of X and using (18), the probability of $\widetilde{x}$ can be calculated as

$$\begin{aligned} P(\widetilde{x}) = \mathbb{E}[\mu_{\widetilde{x}}(X)] &= \int_a^b \frac{x-a}{b-a}g(x)dx + \int_b^c g(x)dx + \int_c^d \frac{d-x}{d-c}g(x)dx \\ &= \frac{1}{b-a}\int_a^b xg(x)dx - \frac{a}{b-a}\left(\Phi(b^*) - \Phi(a^*)\right) + \Phi(c^*) - \Phi(b^*) \\ &\quad + \frac{d}{d-c}\left(\Phi(d^*) - \Phi(c^*)\right) - \frac{1}{d-c}\int_c^d xg(x)dx, \end{aligned} \tag{21}$$

where Φ denotes the c.d.f. of the standard normal distribution, and x^* denotes $(x - m)/\sigma$ for all x. It is easy to show that

$$\int_a^b xg(x)dx = \frac{\sigma}{\sqrt{2\pi}}\left[\exp\left(-\frac{a^{*2}}{2}\right) - \exp\left(-\frac{b^{*2}}{2}\right)\right] + m\left(\Phi(b^*) - \Phi(a^*)\right), \tag{22}$$

which makes it possible to complete the calculation of $P(\widetilde{x})$ using (21).

Let us now compute the expectation of X given $\widetilde{x}$, using the expression (20) for the conditional density of X. We have

$$\mathbb{E}(X|\widetilde{x}) = \frac{\int \mu_{\widetilde{x}}(x)\, x\, g(x)dx}{P(\widetilde{x})}, \tag{23}$$

where the denominator is given by (21). The numerator is

$$\begin{aligned}\int \mu_{\widetilde{x}}(x) x g(x)dx &= \int_a^b \frac{x-a}{b-a}\, x\, g(x)dx + \int_b^c x\, g(x)dx + \int_c^d \frac{d-x}{d-c}\, x\, g(x)dx \\ &= \frac{1}{b-a}\int_a^b x^2\, g(x)dx - \frac{a}{b-a}\int_a^b x\, g(x)dx + \int_b^c x\, g(x)dx \\ &\quad + \frac{d}{d-c}\int_c^d x\, g(x)dx - \frac{1}{d-c}\int_c^d x^2\, g(x)dx, \end{aligned} \tag{24}$$

which can be computed using (22) and

$$\begin{aligned}\int_a^b x^2 g(x)dx &= \frac{\sigma^2}{\sqrt{2\pi}}\left[a^* \exp\left(-\frac{a^{*2}}{2}\right) - b^* \exp\left(-\frac{b^{*2}}{2}\right)\right] \\ &\quad + \frac{2\sigma m}{\sqrt{2\pi}}\left[\exp\left(-\frac{a^{*2}}{2}\right) - \exp\left(-\frac{b^{*2}}{2}\right)\right] \\ &\quad + (m^2+\sigma^2)\left(\Phi(b^*) - \Phi(a^*)\right). \end{aligned} \tag{25}$$

Finally, let us compute

$$\mathbb{E}(X^2|\widetilde{x}) = \frac{\int \mu_{\widetilde{x}}(x)\, x^2\, g(x)dx}{P(\widetilde{x})}. \tag{26}$$

The numerator is

$$\begin{aligned}\int \mu_{\widetilde{x}}(x) x^2 g(x)dx &= \int_a^b \frac{x-a}{b-a}\, x^2\, g(x)dx + \int_b^c x^2\, g(x)dx + \int_c^d \frac{d-x}{d-c}\, x^2\, g(x)dx \\ &= \frac{1}{b-a}\int_a^b x^3\, g(x)dx - \frac{a}{b-a}\int_a^b x^2\, g(x)dx + \int_b^c x^2\, g(x)dx \\ &\quad + \frac{d}{d-c}\int_c^d x^2\, g(x)dx - \frac{1}{d-c}\int_c^d x^3\, g(x)dx, \end{aligned} \tag{27}$$

which can be computed using (25) and

$$\begin{aligned}\int_a^b x^3 g(x)dx &= \frac{\sigma^3}{\sqrt{2\pi}}\left[(2+a^*)\exp\left(-\frac{a^{*2}}{2}\right) - (2+b^*)\exp\left(-\frac{b^{*2}}{2}\right)\right] \\ &\quad + \frac{3\sigma^2 m}{\sqrt{2\pi}}\left[a^* \exp\left(-\frac{a^{*2}}{2}\right) - b^* \exp\left(-\frac{b^{*2}}{2}\right) + \sqrt{2\pi}\left(\Phi(b^*) - \Phi(a^*)\right)\right] \\ &\quad + \frac{3\sigma m^2}{\sqrt{2\pi}}\left[\exp\left(-\frac{a^{*2}}{2}\right) - \exp\left(-\frac{b^{*2}}{2}\right)\right] \\ &\quad + m^3\left(\Phi(b^*) - \Phi(a^*)\right). \end{aligned} \tag{28}$$

References

[1] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):719–725, 2000.

[2] A. Colubi. Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data. *Fuzzy Sets and Systems*, 160(3):344–356, 2009.

[3] E. Côme, L. Oukhellou, T. Denœux, and P. Aknin. Learning from partially supervised data using mixture models and belief functions. *Pattern Recognition*, 42(3):334–348, 2009.

[4] R. Coppi. Management of uncertainty in statistical reasoning: The case of regression. *International Journal of Approximate Reasoning*, 47(3):284–305, 2008.

[5] R. Coppi, M. A. Gil, and H. A. L. Kiers. The fuzzy approach to statistical analysis. *Computational Statistics & Data Analysis*, 51(1):1–14, 2006.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B 39:1–38, 1977.

[7] T. Denœux. Maximum likelihood from evidential data: an extension of the EM algorithm. In C. Borgelt et al., editor, *Combining soft computing and statistical methods in data analysis (Proceedings of SMPS 2010)*, Advances in Intelligent and Soft Computing, pages 181–188, Oviedeo, Spain, 2010. Springer.

[8] T. Denœux, M. Masson, and P.-A. Hébert. Nonparametric rank-based statistics and significance tests for fuzzy data. *Fuzzy Sets and Systems*, 153:1–28, 2005.

[9] T. Denœux and M.-H. Masson. Principal component analysis of fuzzy data using autoassociative neural networks. *IEEE Transactions on Fuzzy Systems*, 12(3):336–349, 2004.

[10] D. Dubois, W. Ostasiewicz, and H. Prade. Fuzzy sets: History and basic notions. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy sets*, pages 21–124. Kluwer Academic Publishers, Boston, 2000.

[11] D. Dubois and H. Prade. *Possibility Theory: An approach to computerized processing of uncertainty*. Plenum Press, New-York, 1988.

[12] P. D'Urso and P. Giordani. A weighted fuzzy c-means clustering model for fuzzy data. *Computational Statistics and Data Analysis*, 50(6):1496–1523, 2006.

[13] J. Gebhardt, M. A. Gil, and R. Kruse. Fuzzy set-theoretic methods in statistics. In R. Slowinski, editor, *Fuzzy sets in decision analysis, operations research and statistics*, pages 311–347. Kluwer Academic Publishers, Boston, 1998.

[14] M. A. Gil and M. R. Casals. An operative extension of the likelihood ratio test from fuzzy data. *Statistical Papers*, 29:191–203, 1988.

[15] M. A. Gil, N. Corral, and M. R. Casals. The likelihood ratio test for goodness of fit with fuzzy experimental observations. *IEEE Transactions on Systems, Man and Cybernetics*, 19(4), 1989.

[16] M. A. Gil, N. Corral, and P. Gil. The minimum inaccuracy estimates in χ^2 tests for goodness of fit with fuzzy observations. *Journal of Statistical Planning and Inference*, 19(1):95–115, 1988.

[17] M. A. Gil, M. López-Díaz, and D. A. Ralescu. Overview on the development of fuzzy random variables. *Fuzzy Sets and Systems*, 157(19):2546–2557, 2006.

[18] P. Giordani and H.A.L. Kiers. Two- and three-way component models for LR fuzzy data in a possibilistic framework. *Fuzzy Sets and Systems*, 157(19):2648–2664, 2006.

[19] G. González-Rodríguez, A. Colubi, P. D'Urso, and M. Montenegro. Multi-sample test-based clustering for fuzzy random variables. *International Journal of Approximate Reasoning*, 50(5):721–731, 2009.

[20] G. González-Rodríguez, A. Colubi, and M. Á. Gil. Fuzzy data treated as functional data: A one-way anova test approach. *Computational Statistics and Data Analysis*, To appear. doi:10.1016/j.csda.2010.06.013.

[21] G. González-Rodríguez, M. Montenegro, A. Colubi, and M. A. Gil. Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. *Fuzzy Sets and Systems*, 157(19):2608–2613, 2006.

[22] H. Hamdan and G. Govaert. Mixture model clustering of uncertain data. In *IEEE International Conference on Fuzzy Systems*, pages 879–884, Reno, Nevada, USA, May 2005. IEEE.

[23] U. Höhle. Masse auf unscharfen Mengen. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36:179–188, 1976.

[24] R. Kruse and K. D. Meyer. *Statistics with vague data*. Kluwer, Dordrecht, 1987.

[25] H. Kwakernaak. Fuzzy random variables. Part I: definitions and theorems. *Inform. Sci.*, 15(1–29), 1978.

[26] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, New York, 1997.

[27] W. Näther. Regression with fuzzy random data. *Computational Statistics & Data Analysis*, 51(1):235–252, 2006.

[28] T. Okuda, H. Tanaka, and K. Asai. A formulation of fuzzy decision problems with fuzzy information using probability mesures of fuzzy events. *Information and Control*, 38:135–147, 1978.

[29] M. L. Puri and D. A. Ralescu. Fuzzy random variables. *J. Math. Anal. Appl.*, 114:409–422, 1986.

[30] B. Quost and T. Denœux. Clustering fuzzy data using the fuzzy EM algorithm. In A. Deshpande and A. Hunter, editors, *Proceedings of the 4th International Conference on Scalable Uncertainty Management (SUM2010)*, number LNAI-6379, pages 333–346, Toulouse, France, September 2010. Springer-Verlag.

[31] G. Schwartz. Estimating the number of components in a finite mixture model. *Annals of Statistics*, 6:461–464, 1978.

[32] R. Viertl. Testing hypotheses with fuzzy data: The fuzzy p-value. *Metrika*, 59(1), 2004.

[33] R. Viertl. Univariate statistical analysis with fuzzy data. *Computational Statistics & Data Analysis*, 51(1):133–147, 2006.

[34] L. A. Zadeh. Fuzzy sets. *Inform. Control*, 8:338–353, 1965.

[35] L. A. Zadeh. Probability measures of fuzzy events. *J. Math. Analysis and Appl.*, 10:421–427, 1968.

[36] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.