Interval-valued EM Algorithm with Application to Estimating Parameters

Zhi-gang Su, Pei-hong Wang, Yi-fan Wang

***Abstract*—This paper investigates on the problem of parameter estimation in statistical model when observations are interval assumed to be related to underlying crisp realizations of a random sample. The proposed approach relies on the extension of likelihood function in interval setting. A maximum likelihood estimate (MLE) of the parameter of interest can then be defined as a crisp value maximizing the generalized likelihood function. Using the Expectation- Maximization (EM) to solve such maximizing problem therefore derives the so-called *interval-valued EM algorithm* (IEM), which makes it possible to solve a wide range of statistical problems involving interval-valued data. As an illustration, the IEM is used to estimate the parameters mean and variance of a univariate normal distribution from interval-valued samples.**

***Index Terms*—EM algorithm, likelihood function, parameter estimation, interval-valued data, univariate normal distribution**

I. INTRODUCTION

In real-life world, there are many kinds of phenomena that are better described by using interval bounds than by suing precise single-valued variables. In fact, intervals take into account the location as well as the variation of the phenomena. Therefore, there emerges a surge of interest in extending the mathematics and theories on precise single-valued variables to imprecise interval-valued variables, for instance, among these, the most popular one is the statistical interval analysis, including interval regression analysis [1]-[8], multidimensional scaling analysis[9], clustering [10]-[11], and so on. However, there are few literatures on parameter estimation from interval-valued data when a parametric statistic model is postulated. Especially, there is not an efficient way used to solve a wide range of statistical problems involving interval-valued data.

Recently, a popular approach, called fuzzy EM algorithm (FEM) [12] and/or evidential EM algorithm (E2M) [13], is proposed to deal with parameter estimation from a postulated parametric statistic model when only fuzzy data and/or uncertain data can be observed. This approach has been successfully applied to a wide range of problems involving fuzzy and/or uncertain data [14]-[15]. The derivations of FEM and E2M implicate that it may be possible to extend the EM algorithm to interval-valued data. With such implication, in this paper, we propose to introduce the likelihood function and EM algorithm to interval-valued data, and then to solve a wide range of statistical problems involving interval-valued data. The proposed approach relies on the extension of likelihood function in interval setting. With the generalized likelihood function, a maximum likelihood estimate (MLE) of the parameter of interest can then be defined as a crisp value maximizing the generalized likelihood functions. Using the Expectation-Maximization (EM) to solve such maximizing problem therefore derives the interval-valued EM algorithm. As will be shown, the interval-valued EM algorithm can be used to solve the problem of parameters estimation in statistical model when observations are interval, through a classical problem.

The rest of this paper is organized as follows. Section 2 and 3 respectively generalizes the likelihood function and EM algorithm to interval-valued data. Section 4 illustrates a classical application of the generalizations in Sections 2 and 3. The last section 5 concludes the paper.

II. LIKELIHOOD FUNCTION IN INTERVAL SETTING

In this section, the generalized likelihood function for interval-valued data is presented. The problem addressed in this paper may be described as follows.

Let $\mathbf{X}$, the *complete-data* vector, be a continuous random vector, taking value in sample space $\Omega_{\mathbf{X}}$ and describing the result of a random experiment. The probability density function (p.d.f.) of $\mathbf{X}$ is denoted by $g(\mathbf{x}, \boldsymbol{\psi})$, where $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_d)'$ is a column vector of unknown parameters with parameter space $\Omega_{\boldsymbol{\psi}}$, where symbol " ' " denotes vector or matrix transposition. Although $\mathbf{X}$ will be generally assumed to be a continuous random vector, $g(\mathbf{x}, \boldsymbol{\psi})$ can still be viewed a probability mass function without confusion in the case where $\mathbf{X}$ is discrete.

If $\mathbf{x}$, a realization of $\mathbf{X}$, is as known exactly, the MLE of $\boldsymbol{\psi}$ as any value maximizing the complete-data likelihood function can be computed

$$L(\boldsymbol{\psi}; \mathbf{x}) = g(\mathbf{x}; \boldsymbol{\psi}) \tag{1}$$

However, the observations $\mathbf{x}$ can be usually imprecise. When $\mathbf{x}$ is not precisely observed but it is known for sure that $\mathbf{x} \in A$ for some (crisp) set $A \subseteq \Omega_{\mathbf{X}}$, we have the following likelihood form [16]:

Manuscript received July 16, 2012, revised August 03, 2012. This work was supported in part by the National Natural Science Foundation of China (No. 51106025).

Zhi-gang Su and Pei-hong Wang are both with the Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education, School of Energy and Environment, Southeast University, Nanjing, 210096, China. (e-mail: Zhigangsu@seu.edu.cn, phwang@seu.edu.cn).

Yi-fan Wang is with the Information Networking Institute. Carnegie Mellon University, Pittsburgh, PA, 15217, USA (e-mail: ivanrex@gmail.com).

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$L(\boldsymbol{\psi};A)=\sum_{\mathbf{x}\in A} g(\mathbf{x};\boldsymbol{\psi}) \tag{2}$$

In a deductive way, when **x** is only imprecise and represented as interval-valued data $I_{\mathbf{x}}$, the likelihood function (2) can be directly generalized by replacing sigma summation with integral as follow:

$$L(\boldsymbol{\psi};I_{\mathbf{x}})=\int_{I_{\mathbf{x}}} g(\mathbf{x};\boldsymbol{\psi})\,d\mathbf{x} \tag{3}$$

To better understand (3), we may firstly understand the interval observation in the following way. The interval observation $I_{\mathbf{x}}$ can be understood as encoding the observer's imperfect knowledge about the realization **x** of a random vector **X**. In this setting, the interval containing **x**, $I_{\mathbf{x}}$, can be interpreted as an interval constraint on the unknown quality **x**. Therefore, the interval $I_{\mathbf{x}}$ can be considered to be generated by a two step process:

1) A realization **x** is drawn from **X**;
2) The observer encodes his/her imperfect knowledge of **x** in the form of a boxcar function, a special form of possibility distribution, defined in Fig. 1.

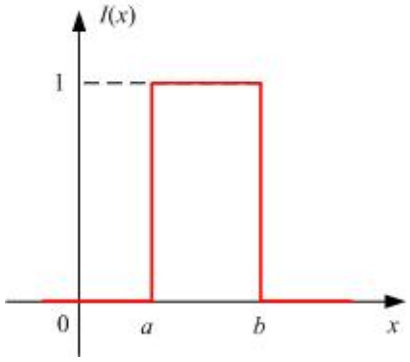


Fig. 1 Boxcar function $I(x)$ for interval $[a, b]$ containing x

It must be stressed that, in the above model, only the first step is considered to be a random experiment. The second step implies gathering information about **x** and modeling this information as a constraint on **x** in the form of a special possibility distribution, and therefore it is not considered as a random experiment.

With above viewpoints, the likelihood function in interval setting can be rewritten as:

$$L(\boldsymbol{\psi};I_{\mathbf{x}})=\int I_{\mathbf{x}}(\mathbf{x})\,g(\mathbf{x};\boldsymbol{\psi})\,d\mathbf{x}=\mathsf{E}_{\boldsymbol{\psi}}\left[I_{\mathbf{x}}(\mathbf{x})\right] \tag{4}$$

Assume that the random vector **X** can be written as $\mathbf{X}=(X_1, X_2, \ldots, X_n)'$, where each X_i is a p-dimensional random vector taking values in $\Omega_{\mathbf{X}}$, and that its realization can be written as $\mathbf{x}=(x_1, x_2, \ldots, x_n)'$, where each realization x_i is imprecisely observed with a interval I_i, obtained by marginalizing the joint interval $I_{\mathbf{x}}$ on X_i. The following two different assumptions can then be made:

1) Under the stochastic *independence* of random variables $X_1, \ldots, X_n$, the probability density function $g(\mathbf{x}, \boldsymbol{\psi})$ can be decomposed as:

$$g(\mathbf{x};\boldsymbol{\psi})=\prod_{i=1}^{n} g(x_i,\boldsymbol{\psi})\mathbb{Z} \tag{5}$$

2) Under the decomposable *assumption* of joint interval $I_{\mathbf{x}}$ containing $\mathbf{x}=(x_1, x_2, \ldots, x_n)'$, we have:

$$I_{\mathbf{x}}(\mathbf{x})=\prod_{i=1}^{n} I_i(x) \tag{6}$$

If both assumptions hold, the likelihood criterion (21) can be written as a product of n items:

$$L(\boldsymbol{\psi};I_{\mathbf{x}})=\prod_{i=1}^{n}\mathsf{E}_{\boldsymbol{\psi}}\left[I_i(X_i)\right] \tag{7}$$

As can be seen from (4) or (7), the likelihood function in interval setting can be viewed as probability of a special fuzzy event $I_{\mathbf{x}}$. In such viewpoint, the likelihood function (4) or (7) is a special case of fuzzy setting.

III. Interval-valued EM Algorithm

With the same notation in Section 2, the EM algorithm [16] approaches the problem of maximizing the observed-data log likelihood $\log L(\psi, A)$ by proceeding iteratively with the complete-data log likelihood $\log L(\psi, \mathbf{x}) = \log g(\mathbf{x}; \psi)$. 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

$$Q_{EM}\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right)=\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left(\log\left[L(\boldsymbol{\psi};\mathbf{x})\right]\middle|A\right) \tag{8}$$

where $\boldsymbol{\psi}^{(q)}$ denotes the current fit of $\boldsymbol{\psi}$ at the q^{th} iteration and $\mathrm{E}_{\psi}^{(q)}(.|A)$ denotes expectation with respect to the conditional distribution of **X** given A, using the parameter vector $\boldsymbol{\psi}^{(q)}$.

The M-steps requires the maximization of $Q(\boldsymbol{\psi}, \boldsymbol{\psi}^{(q)})$ with respect to $\boldsymbol{\psi}$. EM algorithm alternatively repeats the E- and M-steps until the increase of observed data likelihood becomes smaller than some threshold.

The expression (8) can be straightforward extended to interval setting by conditioning on a different probability density. More precisely, the conditional probability density in expectation (8) of $\log L(\psi, \mathbf{x})$ is now replaced by the following probability density function $g(\bullet|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)})$, defined as:

$$g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)=\frac{g\left(\mathbf{x};\boldsymbol{\psi}^{(q)}\right)I_{\mathbf{x}}(\mathbf{x})}{\int I_{\mathbf{x}}(\mathbf{x})\,g\left(\mathbf{x};\boldsymbol{\psi}^{(q)}\right)d\mathbf{x}}=\frac{g\left(\mathbf{x};\boldsymbol{\psi}^{(q)}\right)I_{\mathbf{x}}(\mathbf{x})}{L\left(\boldsymbol{\psi}^{(q)};I_{\mathbf{x}}\right)} \tag{9}$$

The conditional density (9) can be intuitively viewed as the conditional probability of **X** given special fuzzy event $I_{\mathbf{x}}$. At the q^{th} iteration, we therefore have the following computation when only interval $I_{\mathbf{x}}$ can be observed:

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$Q\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right)=\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left(\log\left[L(\boldsymbol{\psi};\mathbf{x})\right]\middle|g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)\right)$$
$$=\frac{\int\log\left[L(\boldsymbol{\psi};\mathbf{x})\right]I_{\mathbf{x}}(\mathbf{x})g\left(\mathbf{x},\boldsymbol{\psi}^{(q)}\right)d\mathbf{x}}{L\left(\boldsymbol{\psi}^{(q)};I_{\mathbf{x}}\right)} \quad (10)$$

We call the EM algorithm using the computation (10) as the E-step interval *EM algorithm*. Finally, the interval EM algorithm also inherits the monotonicity property of the EM algorithm, as shown by the following theorem.

Theorem 1 Any sequence $L(\boldsymbol{\psi}^{(q)};I_{\mathbf{x}})$ for q = 0, 1, 2, … of likelihood values obtained using the interval EM algorithm is nondecreasing, i.e., it verifies

$$L\left(\boldsymbol{\psi}^{(q+1)};I_{\mathbf{x}}\right)\geq L\left(\boldsymbol{\psi}^{(q)};I_{\mathbf{x}}\right)$$

for all q.

Proof

According to (9), we have

$$g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)=\frac{g(\mathbf{x};\boldsymbol{\psi})I_{\mathbf{x}}(\mathbf{x})}{L(\boldsymbol{\psi};I_{\mathbf{x}})}=\frac{L(\boldsymbol{\psi},\mathbf{x})I_{\mathbf{x}}(\mathbf{x})}{L(\boldsymbol{\psi};I_{\mathbf{x}})} \quad (11)$$

Therefore we can define the following expression for $\mathbf{x}\in I_{\mathbf{x}}$:

$$p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)=\frac{L(\boldsymbol{\psi},\mathbf{x})}{L(\boldsymbol{\psi};I_{\mathbf{x}})}=\frac{g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)}{I_{\mathbf{x}}(\mathbf{x})} \quad (12)$$

Taking the log of the leftmost expression of (12), we have

$$\log L(\boldsymbol{\psi};I_{\mathbf{x}})=\log L(\boldsymbol{\psi},\mathbf{x})-\log p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right) \quad (13)$$

Taking the expectation of both sides with respect to $g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)$, we get

$$\log L(\boldsymbol{\psi};I_{\mathbf{x}})$$
$$=\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\log L(\boldsymbol{\psi},\mathbf{x})\middle|g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)\right]-\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\log p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)\middle|g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)\right] \quad (14)$$

By considering the fact that $g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)$ only depends on $I_{\mathbf{x}}$, we rewrite expression (14) as follows:

$$\log L(\boldsymbol{\psi};I_{\mathbf{x}})=\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\log L(\boldsymbol{\psi},\mathbf{x})\middle|I_{\mathbf{x}}\right]-\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\log p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)\middle|I_{\mathbf{x}}\right] \quad (15)$$

$$\log L(\boldsymbol{\psi};I_{\mathbf{x}})=Q\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right)-H\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right) \quad (16)$$

with $H\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right)=\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\log p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)\middle|I_{\mathbf{x}}\right]$.

We thus have:

$$\log L\left(\boldsymbol{\psi}^{(q+1)};I_{\mathbf{x}}\right)-\log L\left(\boldsymbol{\psi}^{(q)};I_{\mathbf{x}}\right)=Q\left(\boldsymbol{\psi}^{(q+1)},\boldsymbol{\psi}^{(q)}\right)-$$
$$Q\left(\boldsymbol{\psi}^{(q)},\boldsymbol{\psi}^{(q)}\right)-\left(H\left(\boldsymbol{\psi}^{(q+1)},\boldsymbol{\psi}^{(q)}\right)-H\left(\boldsymbol{\psi}^{(q)},\boldsymbol{\psi}^{(q)}\right)\right) \quad (17)$$

The first difference on the right-hand side of (17) is nonnegative as $\boldsymbol{\psi}^{(q+1)}$ has been chosen to maximize $Q(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)})$ with respect to parameters $\boldsymbol{\psi}$. It thus remains to check that the second difference on the right-hand side of (17) is non-positive. In other words, we need to verify the following inequality:

$$H\left(\boldsymbol{\psi}^{(q+1)},\boldsymbol{\psi}^{(q)}\right)-H\left(\boldsymbol{\psi}^{(q)},\boldsymbol{\psi}^{(q)}\right)\leq 0 \quad (18)$$

Now for any $\boldsymbol{\psi}$, we have

$$H\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right)-H\left(\boldsymbol{\psi}^{(q)},\boldsymbol{\psi}^{(q)}\right)=\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\log p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)\middle|I_{\mathbf{x}}\right]-$$
$$\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\log p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)\middle|I_{\mathbf{x}}\right]=\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\log\frac{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)}{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)}\middle|I_{\mathbf{x}}\right] \quad (19)$$

According to Jensen's inequality, we get

$$H\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right)-H\left(\boldsymbol{\psi}^{(q)},\boldsymbol{\psi}^{(q)}\right)\leq\log\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\frac{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)}{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)}\middle|I_{\mathbf{x}}\right]=0 \quad (20)$$

because of the following equality:

$$\log\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\frac{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)}{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)}\middle|I_{\mathbf{x}}\right]=\log\mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left[\frac{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)}{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)}\middle|g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)\right]$$
$$=\log\int\frac{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)}{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)}g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)d\mathbf{x}$$
$$=\log\int\frac{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)}{p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)}I_{\mathbf{x}}(\mathbf{x})p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)d\mathbf{x}$$
$$=\log\int p\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)I_{\mathbf{x}}(\mathbf{x})d\mathbf{x}$$
$$=\log\int g\left(\mathbf{x}\middle|I_{\mathbf{x}};\boldsymbol{\psi}\right)d\mathbf{x}=0$$

The proof is thus completed. ■

IV. Application: Estimation for Univariate Normao Distribution from Interva-Valued Data

Assuming that the complete data x is a realization of an independent identical distribution random sample from a normal distribution with mean m and standard deviation σ. The observed data are supposed to be interval $I_{\mathbf{x}}$. The complete-data p.d.f. can therefore be defined as

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$g(\mathbf{x};\boldsymbol{\psi}) = \prod_{i=1}^{n} g(x_i;\boldsymbol{\psi}) \tag{21}$$

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

With (21), the complete-data log likelihood is thus:

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

Taking the expectation of $\log L(\boldsymbol{\psi}, \mathbf{x})$ conditionally on the observed interval $I_{\mathbf{x}}$ and using the fit $\boldsymbol{\psi}^{(q)}$ of $\boldsymbol{\psi}$ to perform the E-step, it can get

$$Q\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right) = \mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left(\log\left[L(\boldsymbol{\psi};\mathbf{x})\right] \middle| g\left(\mathbf{x}\middle| I_{\mathbf{x}};\boldsymbol{\psi}^{(q)}\right)\right) = -\frac{n}{2}\log(2\pi) -$$
$$n\log\sigma - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} \mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left(X_i^2 \middle| I_i(x_i)\right) - 2m\sum_{i=1}^{n} \mathsf{E}_{\boldsymbol{\psi}^{(q)}}\left(X_i \middle| I_i(x_i)\right) + nm^2\right) \tag{22}$$

where $\alpha_i^{(q)} = \mathsf{E}_{\boldsymbol{\psi}^{(q)}}(X_i^2 | I_i(x_i))$ and $\beta_i^{(q)} = \mathsf{E}_{\boldsymbol{\psi}^{(q)}}(X_i | I_i(x_i))$ can be computed using (25)~(28) in Appendix A.

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

$$\frac{\partial Q\left(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}\right)}{\partial m} = -2\sum_{i=1}^{n} \beta_i^{(q)} + 2nm$$

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

Equating these derivatives to zero and solving for **b** and σ, we get the following unique solution:

$$m^{(q+1)} = \frac{1}{n}\sum_{i=1}^{n} \beta_i^{(q)} \tag{23}$$

$$\sigma^{(q+1)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \alpha_i^{q} - \left(m^{(q+1)}\right)^2} \tag{24}$$

Example 1

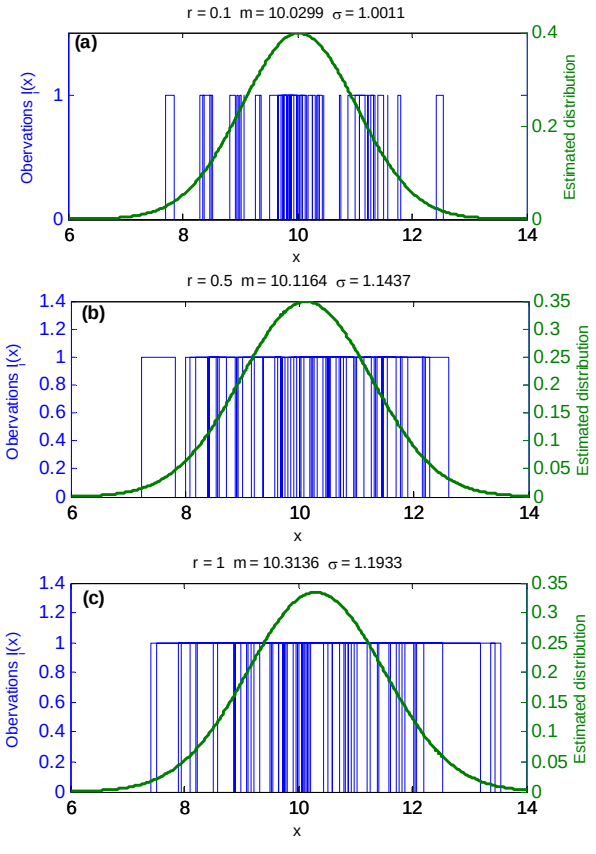
In this example, we aim to estimate the parameters (i.e., the mean and standard deviation) of univariate normal distribution when only interval observations can be obtained. The interval observations are generated as follows. Suppose n realizations x_i (i = 1, 2, …, n) are drawn from the normal univariate distribution with mean m and variance σ, and therefore the number of n interval observations $I_{\mathbf{x}} = \{I_1(x) I_2(x), \ldots, I_n(x)\}$ are constructed as: $I_i(x) = 1$ when $x \in [x_i - r_i, x_i + r_i]$, otherwise $I_i(x) = 0$, where the half bandwidth r_i is randomly generated from the interval $[0, r]$. In addition, we aim to

1) see the way how the half bandwidth r influence on the estimation, given the number n, mean m and variance σ;
2) see the whether the observation number n affect the estimation or not; given the half bandwidth r, mean m, and variance σ.

For convenience, we consider the following four cases for r: $r = \{0.1, 0.5, 1, 1.5\}$ and three cases for n: $n = \{10, 20, 40\}$, in the condition mean m = 10 and variance σ = 1 (Note that, one can consider other mean and variance as the parameters to be estimated.). In each experiment, the sample mean and variance computed over the centers of interval observations are taken as the initial estimate of m and σ.

- *Case 1) study*

In this case study, we suppose n = 40. (AS will be shown, the bigger n is, the smaller estimation error is). For each case $r \in \{0.1, 0.5, 1, 1.5\}$, the experiment is implemented 100 times and one of which is randomly selected to be illustrated. Therefore, total four of the four groups of 100 trials can be obtained, shown in Fig. 2. Correspondingly, Fig. 3 presents the box plot of estimations of the mean and deviation over the four groups of 100 trials.



ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

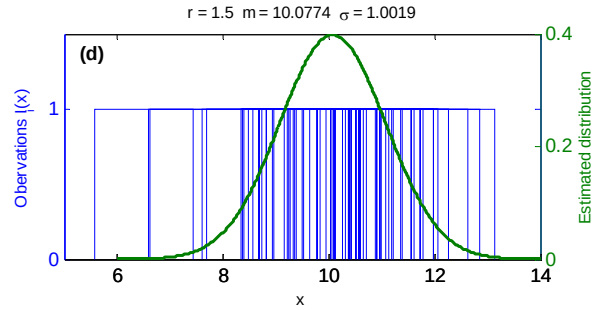


Fig. 2 Estimated distributions of univiate normal with mean m = 10 and deviation σ= 1 in different cases: (a) r = 0.1, (b) r = 0.5, (c) r = 1, (d) r = 1.5.

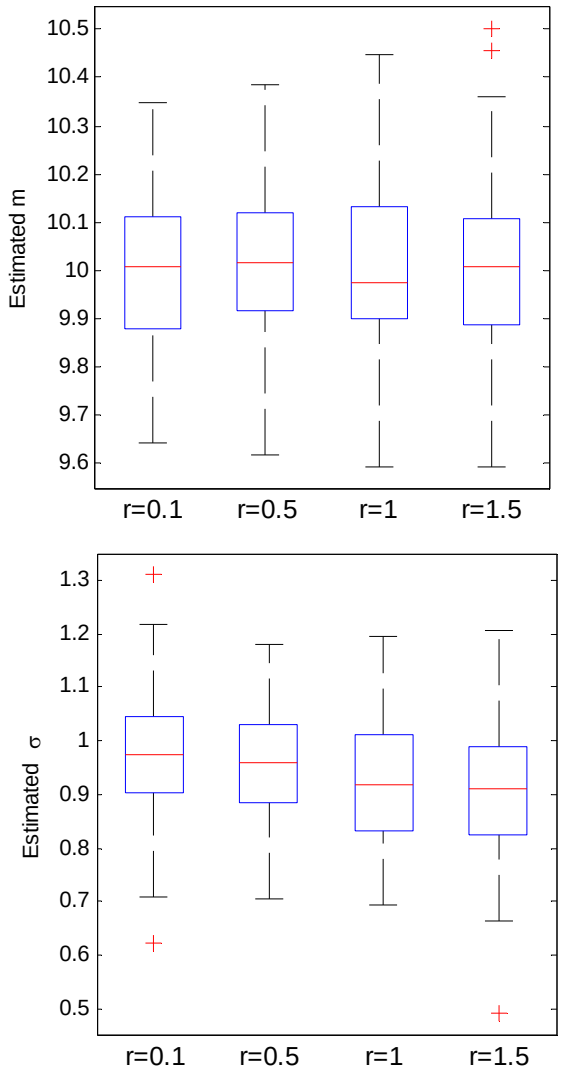


Fig. 3 The estimated mean m and variance σ in the cases r = 0.1, 0.5, 1 and 1.5 when n =40

As can be seen from Figs. 2 and 3, the estimated mean and deviation approach to the true ones when different length of interval observations can be obtained, and the estimated parameters becomes more accurate with decreasing of the width of the interval observations.

- *Case 2) study*

In this case study, we suppose r = 0.1. (AS can be seen from case 1) study, when r takes small value, the estimation result is more accurate.). For each case $n \in \{10, 20, 40\}$, the experiment is implemented 100 times. The estimation results for mean and variance are presented in Fig. 4. From Fig. 4, we can see that, the larger the observation number n is, the estimation accurate is much higher.

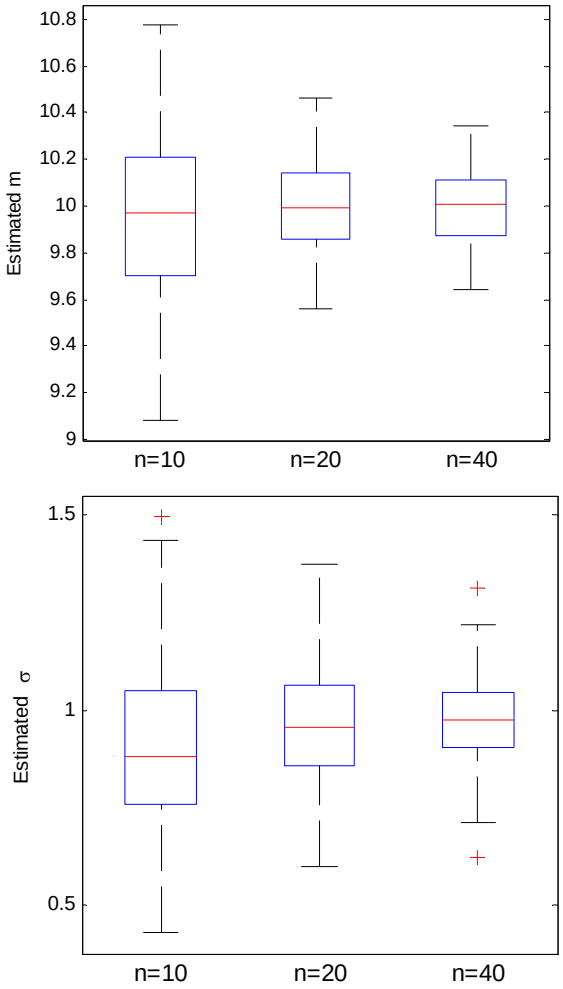


Fig. 4 The estimated mean m and variance σ in the cases n = 10, 20 and 40 when r = 0.1

From both the cases 1) and 2) studies, we can see that our proposed method can be used to estimate parameters involving interval-valued data, and we obtain the similar results as those done in the classical maximum likelihood estimation by using EM algorithm: the larger the observation number is, the the smaller the estimation error is, and the more precise the observations are, the smaller the estimation error is (i.e., the smaller the half bandwidth r is, the smaller the estimation accurate is).

V. CONCLUSIONS

This paper proposes an approach to solve the problem of parameters estimation in statistical model involving interval-valued data. In this approach, the interval-valued data is understood as the observer encoding his/her imperfect knowledge in the form of boxcar function on a realization of a random vector. With such understanding, the likelihood function is extended to interval setting, and then a maximum likelihood estimate of the parameter of interest is defined as a crisp value maximizing the generalized likelihood function. Such maximizing problem is achieved by using the interval-valued EM algorithm, which is an extension of the

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

classical EM algorithm.

As an illustration, the proposed approach is used to estimate the mean and variance of univariate normal distribution. From the illustration result, we have reason to believe that it is possible to apply interval-valued EM algorithm to solve a wide range of statistical problems involving interval-valued data.

In further, it is interesting to apply the interval-valued EM algorithm to solve different problems involving interval-valued data, for instance, to estimate the regression coefficients of linear and nonlinear regression model involving interval-valued data, so as to establish regression model with crisp inputs and interval output, which is useful in engineering application where only interval values can be obtained for predicting a process parameter.

APPENDIX

Suppose the interval observation $I_i(x) = [a_i, b_i]$ and the p.d.f. $g(x)$ with parameters $\boldsymbol{\psi} = (m, \sigma)'$, we have:

$$\mathsf{E}_{\boldsymbol{\psi}}\left(X_i \middle| I_i(x)\right) = \frac{\int x g(x) I_i(x)\,dx}{\int g(x) I_i(x)\,dx} = \frac{\int x g(x) I_i(x)\,dx}{L\left(\boldsymbol{\psi}, I_i(x)\right)} \tag{25}$$

where the denominator is given by (4). The numerator is

$$\begin{aligned}\int x g(x) I_i(x)\,dx &= \int_{a_i}^{b_i} x g(x)\,dx \\ &= \frac{\sigma}{\sqrt{2\pi}}\left[\exp\left(-\frac{a_i^{*2}}{2}\right) - \exp\left(-\frac{b_i^{*2}}{2}\right)\right] + m\left(\Phi\left(b_i^*\right) - \Phi\left(a_i^*\right)\right)\end{aligned} \tag{26}$$

where $\Phi(.)$ denotes the c.d.f. of the standard normal distribution, and x^* denotes $(x-m)/\sigma$ for all x.

We finally compute

$$\mathsf{E}_{\boldsymbol{\psi}}\left(X_i^2 \middle| I_i(x)\right) = \frac{\int x^2 g(x) I_i(x)\,dx}{\int g(x) I_i(x)\,dx} = \frac{\int x^2 g(x) I_i(x)\,dx}{L\left(\boldsymbol{\psi}, I_i(x)\right)} \tag{27}$$

with

$$\begin{aligned}\int x^2 g(x) I_i(x)\,dx &= \int_{a_i}^{b_i} x^2 g(x)\,dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\left[a_i^* \exp\left(-\frac{a_i^{*2}}{2}\right) - b_i^* \exp\left(-\frac{b_i^{*2}}{2}\right)\right] \\ &+ \frac{2\sigma m}{\sqrt{2\pi}}\left[a_i^* \exp\left(-\frac{a_i^{*2}}{2}\right) - b_i^* \exp\left(-\frac{b_i^{*2}}{2}\right)\right] \\ &+ \left(m^2 + \sigma^2\right)\left(\Phi\left(b_i^*\right) - \Phi\left(a_i^*\right)\right)\end{aligned} \tag{28}$$

ACKNOWLEDGMENT

The authors are grateful to the contributions of Prof. Thierry Denoeux for our work.

REFERENCES

[1] H. Tanaka, "Fuzzy data analysis by possibilistic linear models," *Fuzzy Sets and Systems*, vol. 24, no. 3, pp. 363–375, Dec. 1987.
[2] P.Y. Hao, "Interval regression analysis using support vector networks," *Fuzzy sets and systems*, vol. 160, no. 17, pp. 2466-2485, Sep. 2009.
[3] M. Hladik and M. Cerny, "Interval regression by tolerance analysis approach," *Fuzzy sets and systems*, vol. 193, pp. 85-107, Apr. 2012.
[4] Y.-C. Hu, "Function-link nets with genetic algorithm based learning for robust nonlinear interval regression analysis," *Neurocomputing*, vol.72, no. 7-9, pp.1808-1816, Mar. 2009.
[5] C. Hwang, D.H. Hong, and K.H. Seok, "Support vector interval regression machine for crisp input and output data," *Fuzzy sets and systems*, vol. 157, no. 8, pp. 1114-1125, Apr. 2006.
[6] H. Ishibuchi and H. Tanaka, "Several formulations of interval regression analysis". *Proc. Sino-Japan Joint Meeting on Fuzzy Sets and Systems*, Beijing, China, 1990, pp. B2–2.
[7] J.-S. Jeng, C.-C. Chuang, and S.-F. Su, "Support vector interval regression networks for interval regression analysis," Fuzzy sets and systems, vol.138, no. 2, pp. 283-300, Sep. 2003.
[8] M. Sato-Ilic, "Symbolic clustering with interval-valued data," *Procedia Computer Science*, vol. 6, pp. 358-263, 2011.
[9] T. Denoeux and M. Masson, "Multidimensional scaling of interval-valued dissimilarity data," *Pattern recognition letters*, vol. 21, no.1, pp. 83-92, Jan. 2000.
[10] F. Palumbo and C.N. Lauro, "A PCA for interval-valued data based on midpoints and radii". In: *new developments in Psychometrics*, H. Yanai, A. Okada, K. Shigemasu, Y. Kano and J. Meulman, Eds., *Psychometric Society*, 2003, pp. 641-648.
[11] H. Wang, R. Guan, and J. Wu, "CIPCA: Complete-information- based principal component analysis for interval-valued data," *Neurocomputing*, vol. 86, pp.158-169, Jun. 2012.
[12] T. Denoeux, "Maximum likelihood from fuzzy data using the EM algorithm," *Fuzzy sets and systems*, vol. 183, no. 1, pp. 72-91, Nov. 2011.
[13] T. Denoeux, "Maximum likelihood estimation from uncertain data in the belief function framework," *IEEE Transactions on knowledge and data engineering*, to be published.
[14] B. Quost and T. Denoeux, "Clustering fuzzy data using the fuzzy EM algorithm," In: *the 4th international conference on Scalable uncertainty management*, Toulouse, France, 2010, pp. 333-346.
[15] Z.-G. Su, P.-H. Wang, and Z.-L. Song, "Kernel based nonlinear fuzzy regression model", *Engineering applications of Artificial Intelligence*, to be published.
[16] 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*, vol. 39, pp.1–38, 1977.