RESEARCH

Open Access

CrossMark

Percentile Matching Estimation of Uncertainty Distribution

S Sampath[*] and K. Anjana

* Correspondence:
sampath1959@yahoo.com
Department of Statistics, University
of Madras, Chennai 600005, India

Abstract

This paper considers the application of method of percentile matching available in statistical theory of estimation for estimating the parameters involved in uncertainty distributions. An empirical study has been carried out to compare the performance of the proposed method with the method of moments and the method of least squares considered by Wang and Peng (J. Uncertainty Analys. Appl. **2**, (2014)) and Liu (Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty, (2010)), respectively. The numerical study clearly establishes the superiority of the proposed method over the other two methods in estimating the parameters involved in linear uncertainty distribution when appropriate orders of percentiles are used in the estimation process.

Keywords: Uncertainty distribution, Uncertain statistics, Method of least squares, Method of moments, Method of percentile matching

Introduction

Indeterminacy that occurs in real-life situations when the outcome of a particular event is unpredictable in advance leads to uncertainty. According to Liu [5], frequency generated from samples (historical data) and belief degree evaluated by domain experts are the two ways to explain indeterminate quantities. A fundamental premise of axiomatic approach of probability theory which came into existence in 1933 is that the estimated probability distribution should be close to the long run cumulative frequency. This approach is reliable when large samples are available. In cases where samples are not available for estimating the unknown parameters of uncertainty distributions, the only choice left out is to go for belief degrees. Belief degree refers to the belief of individuals on the occurrence of events. In order to model belief degrees, Liu [2] introduced the uncertainty theory. Since then, it has developed vigorously throughout the years. Liu [4] explains the need for uncertainty theory. Zhang [8] discusses about characteristics of uncertain measure. Liu [5] explains linear, zigzag, normal, lognormal, and empirical uncertainty distributions.

Several methods are available for estimating the unknown parameters of probability distributions. Method of least squares, method of moments, and method of maximum likelihood are some among them. Method of moments is one of the popular methods meant for estimating parameters in a probability distribution. Method of maximum likelihood is an equally popular estimation method possessing several optimum properties. Method of least squares is a common technique mainly used for estimating parameters of regression models.

Analogous to various methods used in probability theory, estimation techniques have also been developed in uncertainty theory. Uncertain statistics refers to a methodology used for collecting and interpreting expert's experimental data by uncertainty theory. The study of uncertain statistics was started by Liu [3]. Liu [2, 5] introduced the concept of moments in uncertainty theory. Wang and Peng [7] proposed the method of moments as a technique for estimating the unknown parameters of uncertainty distributions. Liu [5] gives detailed explanation of method of least squares, method of moments, and Delphi method. Apart from these methods, exploration on the applications of alternative methods remains unattended. Method of percentile matching is an estimation technique used in statistical theory of estimation which plays a vital role in dealing with estimation of parameters when other popular methods fail to be effective. More details about the method of percentile matching can be found in Klugman et al. [1]. The absence of concepts like uncertainty density function makes the task of defining a function similar to likelihood function (available in statistical theory) a difficult one. Hence, adopting a method similar to the maximum likelihood estimation in the uncertainty framework becomes difficult. In this paper, it is proposed to investigate the utility of the method of percentile matching in estimating the unknown parameters of uncertainty distributions. It is proposed to compare the percentile matching method with the existing competitors by way of numerical studies.

The paper is organized as follows. The second section of this paper gives a detailed description on preliminary concepts of uncertainty theory. The third section deals with the commonly used estimation methods in probability theory and also explains the method of percentile matching. The fourth section is devoted for discussion on methods meant for estimating unknown parameters of uncertainty distributions. The fifth section discusses the experimental studies carried out for estimating the unknown parameters of linear uncertainty distribution. Findings and conclusions are given in the sixth section.

Uncertainty Theory

This section gives a short description on different terminologies associated with uncertainty theory due to Liu [2, 3, 5]. Certain results due to Sheng and Kar [6] on linear uncertainty distribution and details about the method developed by Wang and Peng [7] for estimating parameters of uncertainty distributions are also discussed.

Let Γ be a nonempty set and $\mathcal{L}$ be a σ-algebra over Γ. Each element $\Lambda \in \mathcal{L}$ is called an event. A number $\mathcal{M}(\Lambda)$ indicates the level that Λ will occur.

Uncertain measure: Liu [2] defines a set function $\mathcal{M}$ to be an uncertain measure if it satisfies the following three axioms:

Axiom 1 (normality axiom) $\mathcal{M}\{\Gamma\} = 1$.

Axiom 2 (duality axiom) $\mathcal{M}\{\Lambda\} + \mathcal{M}\{\Lambda^c\} = 1$.

Axiom 3 (subadditivity axiom) For every countable sequence of events,

$$\mathcal{M}\left\{ \bigcup_{i=1}^{\infty} \Lambda_i \right\} \leq \sum_{i=1}^{\infty} \mathcal{M}\{\Lambda_i\}.$$

Although the probability measure satisfies the first three axioms, the probability theory is not a special case of uncertainty theory because product probability measure does not satisfy the product axiom.

Axiom 4 (product axiom) Let $(\Gamma_k, \mathcal{L}_k, \mathcal{M}_k)$ be uncertainty spaces for $k = 1, 2, 3, \ldots$. The product uncertain measure is an uncertain measure satisfying

$$\mathcal{M}\left\{\prod_{k=1}^{\infty} \Lambda_k\right\} = \bigwedge_{k=1}^{\infty} \mathcal{M}_k\{\Lambda_k\}$$

where Λ_k are arbitrarily chosen events from $\mathcal{L}_k$ for $k = 1,2,3,\ldots$, respectively.

Uncertain variable: It is defined by Liu [2] as a measurable function ξ from an uncertainty space $(\Gamma, \mathcal{L}, \mathcal{M})$ to the set of real numbers such that $\{\xi \in B\}$ is an event for any Borel set B of real numbers.

Uncertainty distribution: Uncertainty distribution Φ of an uncertain variable ξ is defined by Liu [2] as

$\Phi(x) = \mathcal{M}\{\xi \le x\}, \forall x \in \mathfrak{R}.$

Liu [5] made studies on various uncertainty distributions, namely, linear, zigzag, normal, and lognormal. This work is related to linear uncertainty distribution stated below.

Linear uncertainty distribution: An uncertain variable ξ is called linear if it has uncertainty distribution of the form

$$\Phi(x) = \begin{cases} 0, & if \ x \le a \\ \frac{x-a}{b-a}, & if \ a \le x \le b \\ 1, & if \ x \ge b \end{cases}$$

where a and b are real numbers with $a < b$. It is usually denoted by $\mathcal{L}(a,b)$.

Empirical uncertainty distribution (Liu [3]): Empirical uncertainty distribution based on a given experimental data is defined as

$$\Phi_n(x) = \begin{cases} 0, & if \ x < x_1 \\ \alpha_i + \frac{(\alpha_{i+1}-\alpha_i)(x-x_i)}{x_{i+1}-x_i}, & if \ x_i \le x \le x_{i+1}, 1 \le i < n \\ 1, & if \ x > x_n \end{cases}$$

where $x_1 < x_2 < \ldots < x_n$ and $0 \le a_1 \le a_2 \le \ldots \le a_n \le 1$.

Regular uncertainty distribution: An uncertainty distribution $\Phi(x)$ is said to be regular by Liu [3] if it is a continuous and strictly increasing function with respect to x where $0 < \Phi(x) < 1$ and $\lim_{x \to -\infty} \Phi(x) = 0$, $\lim_{x \to +\infty} \Phi(x) = 1$. For example, linear, zigzag, normal, and lognormal uncertainty distributions are all regular.

Expected value of an uncertain variable: The expected value of an uncertain variable ξ is defined by Liu [2] as

$$E[\xi] = \int_0^{+\infty} \mathcal{M}\{\xi \ge x\}\,dx - \int_{-\infty}^0 \mathcal{M}\{\xi \le x\}\,dx$$

provided that at least one of the two integrals is finite.

It has been shown by Liu [2] that

$$E[\xi] = \int_0^{+\infty} (1-\Phi(x))\,dx - \int_{-\infty}^0 \Phi(x)\,dx.$$

Also from this expression, using integration by parts Liu [3] gets

$$E[\xi] = \int_{-\infty}^{+\infty} x \, d\,\Phi(x).$$

If ξ has a regular uncertainty distribution Φ, then by substituting $\Phi(x)$ with a, x with $\Phi^{-1}(\alpha)$ in the previous expression and following the change of variables of integral, Liu [3] gives

$$E[\xi] = \int_{0}^{1} \Phi^{-1}(\alpha) \, d\alpha.$$

Moments: If ξ is an uncertain variable and k is a positive integer, then Liu [2] gives the kth moment of ξ as $E[\xi^k]$.

Let ξ be an uncertain variable with uncertainty distribution Φ. Then by Liu [5],

(i) If k is an odd number, then the kth moment of ξ is defined as

$$E\left[\xi^k\right] = \int_{0}^{+\infty} \left(1 - \Phi\left(\sqrt[k]{x}\right)\right) dx - \int_{-\infty}^{0} \Phi\left(\sqrt[k]{x}\right) dx.$$

(ii) If k is an even number, then the kth moment of ξ is defined as

$$E\left[\xi^k\right] = \int_{0}^{+\infty} \left(1 - \Phi\left(\sqrt[k]{x}\right) + \Phi\left(-\sqrt[k]{x}\right)\right) dx.$$

(iii) If k is a positive integer, then the kth moment of ξ is defined as

$$E[\xi^k] = \int_{-\infty}^{+\infty} x^k \, d\,\Phi(x).$$

Sheng and Kar [6] proved that, if an uncertain variable ξ has a regular uncertainty distribution Φ and k is a positive integer, then the kth moment of ξ is

$$E\left[\xi^k\right] = \int_{0}^{1} \left(\Phi^{-1}(\alpha)\right)^k d\alpha.$$

Sheng and Kar [6] derived the expressions for the first three moments of a linear uncertain variable $\xi \sim \mathcal{L}(a, b)$. They are given below.

$$E[\xi] = \frac{a + b}{2}, \tag{1}$$

$$E[\xi^2] = \frac{a^2 + ab + b^2}{3}, \tag{2}$$

$$E[\xi^3] = \frac{(a + b)(a^2 + b^2)}{4}. \tag{3}$$

For a given expert's experimental data

$$(x_1, \alpha_1), (x_2, \alpha_2), ..., (x_n, \alpha_n)$$

that meet the condition

$$0 \le x_1 < x_2 < ... < x_n, \, 0 \le \alpha_1 \le \alpha_2 \le ... \le \alpha_n \le 1$$

where $x_r's$ are the observed values and $\alpha_r's$ are the respective belief degree values, Wang and Peng [7] gives the kth empirical moment of the uncertain variable ξ based on empirical uncertainty distribution as

$$\overline{\xi}^k = \alpha_1 x_1{}^k + \frac{1}{k+1} \sum_{i=1}^{n-1} \sum_{j=0}^{k} (\alpha_{i+1} - \alpha_i) \, x_i^j x_{i+1}{}^{k-j} + (1 - \alpha_n) x_n{}^k. \tag{4}$$

Thus, in this section, prerequisites needed for tackling the main problem considered in this paper have been presented. Before venturing to the problem of estimating the parameters of an uncertainty distribution, some of the common methods employed in estimating the unknown parameters of probability distributions are briefly listed in the following section.

Methods for Estimating the Unknown Parameters of Probability Distributions

Some of the frequently used methods for estimating the parameters of probability distributions include method of moments, method of maximum likelihood, and method of least squares. These methods can be used for any probability distributions. However, their efficiencies will depend on the nature of distributions. Three methods of estimation, namely, method of moments, method of least squares, and the method of percentile matching, are presented below. Method of least squares is used for estimating parameters by minimizing the squared distance between the observed data and their fitted values. Method of moments is possibly the oldest method of finding point estimators. The moment estimators are obtained by equating the first k sample moments to the corresponding k population moments and solving the system of resulting simultaneous parametric equations in terms of sample moments. Percentile matching method uses percentiles of different orders of available data towards estimation of parameters in a distribution. When sample data is discrete, finding the percentile involves smoothing of data. If there are n observations, then the kth percentile is found by interpolating between the two data points that are around the nkth observation. A percentile matching estimate of the vector valued parameter $\theta = (\theta_1, \theta_2, ..., \theta_p)$ is any solution of the p equations $\pi_{g_k}(\theta) = \hat{\pi}_{g_k}, k = 1, 2, 3, ..., p$ where $g_1, g_2, ... g_p$ are p arbitrarily chosen percentiles. Here, the left hand side, namely, $\pi_{g_k}(\theta)$, represents the theoretical percentile of order $g_k (k = 1, 2, ..., p)$ and the right hand side $\hat{\pi}_{g_k}$ gives the corresponding smoothed empirical estimate. It is pertinent to note that the smoothed empirical estimate of a percentile is found by $\hat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}$ where $j = \lfloor (n+1)g \rfloor$ and $h = (n+1)g - j$. $\lfloor . \rfloor$ indicates the greatest integer function and $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$ are the order statistics from the sample.

It is to be noted that $F(\pi_{g_k} | \theta) = g_k, k = 1, 2, 3, ..., p$ where $F(.)$ is the true cumulative distribution of the underlying probability distribution. On substituting $\hat{\pi}_{g_k}$ in lieu of π_{g_k} in this equation, a system of p equations is formed. Solutions based on this system of equations are taken as estimates of the parameters. It may be noted that $\hat{\pi}_g$ cannot be obtained for $g < \frac{1}{(n+1)}$ or $g > \frac{n}{(n+1)}$. This is reasonable as it is not meaningful to find the value of very large or small percentiles from small samples. Smoothed version is used whenever an empirical percentile estimate is needed.

The next section will explain the method of moments for estimating the unknown parameters of linear uncertainty distribution.

Methods for Estimating Parameters of Uncertainty Distributions

The problem of estimating parameters involved in uncertainty distributions has received the attention of researchers. Some methods parallel to those available in

statistical estimation theory have been utilized by uncertainty researchers. Two methods used in uncertain parameter estimation, namely, method of least squares and method of moments, are explained below.

Method of Least Squares

The method of least squares is due to Liu [3]. Suppose that an uncertainty distribution to be determined has a known functional form $\Phi(x|\theta_1, \theta_2, ..., \theta_p)$ having parameters $\theta_1, \theta_2, ..., \theta_p$. To estimate the parameters $\theta_1, \theta_2, ..., \theta_p$, the method of least squares minimizes the sum of the squares of the distance of expert's experimental data from the uncertainty distribution. For a given set of expert's experimental data

$$(x_1, \alpha_1), (x_2, \alpha_2), .., (x_n, \alpha_n),$$

the least square estimates of $\theta_1, \theta_2, ..., \theta_p$ are found by minimizing

$$\sum_{i=1}^{n} \left(\Phi\left(x_i | \theta_1, \theta_2, ..., \theta_p\right) - \alpha_i \right)^2$$

with respect to θ.

While estimating uncertain parameters, it may be noted that closed form solutions for least square estimates may not be available always. Hence most of the times certain tools available in numerical mathematics are employed to estimate such parameters. A MATLAB toolbox available in literature makes the computation of least square estimates an easy one.

Method of Moments

Wang and Peng [7] proposed the method of moments for estimating unknown parameters of an uncertainty distribution. The method is as follows.

Let a non-negative uncertain variable ξ have an uncertainty distribution $\Phi(x|\theta_1, \theta_2, ..., \theta_p)$ with unknown parameters $(\theta_1, \theta_2, ..., \theta_p)$. Given a set of expert's experimental data

$$(x_1, \alpha_1), (x_2, \alpha_2), .., (x_n, \alpha_n)$$

where $0 \leq x_1 < x_2 < ... < x_n, 0 \leq \alpha_1 \leq \alpha_2 \leq ... \leq \alpha_n \leq 1$.

The expression for kth empirical moment of ξ due to Wang and Peng [7] obtained with the help of empirical uncertainty distribution is given by (4). The moment estimates $\left(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p\right)$ are obtained by equating the first p theoretical moments of ξ to the corresponding empirical moments. That is, the moment estimates should solve the system of equations,

$$\int_{0}^{+\infty} \left(1 - \Phi\left(\sqrt[k]{x}|\theta_1, \theta_2, ..., \theta_p\right)\right) dx = \overline{\xi}^k, \ k = 1, 2, 3, ..., p$$

where $\overline{\xi}^1, \overline{\xi}^2, ..., \overline{\xi}^p$ are the empirical moments found using (4).

For example, let ξ be a linear uncertain variable $\xi \sim \mathcal{L}(a, b)$ with two unknown parameters a and b which are two positive real numbers satisfying $a < b$. The linear uncertainty distribution function is given by

$$\Phi(x) = \begin{cases} 0, & if \ x \le a \\ \dfrac{x-a}{b-a}, & if \ a \le x \le b \\ 1, & if \ x \ge b. \end{cases}$$

Since there are two unknown parameters in a linear uncertainty distribution, method of moments makes use of the first and second theoretical and empirical moments of a linear uncertain variable. First and second theoretical moments of a linear uncertain variable derived by Sheng and Kar [6] are given by (1) and (2), respectively. First and second empirical moments denoted by $\overline{\xi}^1$ and $\overline{\xi}^2$ are calculated using the expression given in (4) by putting $k = 1$ and 2, respectively. Equating the first and second theoretical moments to the corresponding empirical moments and solving the resulting quadratic equation, the estimates of unknown parameters are obtained. Minimum among the positive and negative roots give $\hat{a}$ and the maximum among them gives $\hat{b}$.

In this paper, the utility of percentile matching method for estimation of parameters of uncertainty distributions is examined.

Method of Percentile Matching

Given a set of expert's experimental data

$$(x_1, \alpha_1), (x_2, \alpha_2), .., (x_n, \alpha_n)$$

where $x_i, i = 1, 2, 3, ..., n$ are the observed values and $\alpha_i, i = 1, 2, 3, ..., n$ are the belief degrees. Here, it is assumed that $0 \le x_1 < x_2 < ... < x_n, 0 \le \alpha_1 \le \alpha_2 \le ... \le \alpha_n \le 1$.

The observed values $x_i, i = 1, 2, 3, ..., n$ are expected to lie in the interval (*a*,*b*). Following the definition given by Liu [3] which was stated in the "Uncertainty Theory" section, empirical uncertainty distribution can be constructed. An empirical percentile of order k of an uncertainty distribution is defined as the solution of the equation $\Phi_n(x) = \frac{k}{100}$ where $\Phi_n(x)$ is the smoothed empirical uncertainty distribution. Similarly, a theoretical percentile of order k is defined as the solution of the equation $\Phi(x) = \frac{k}{100}$ where $\Phi(x)$ is the true uncertainty distribution. As in the case of percentile matching method meant for probability distributions, p empirical percentiles of desired orders are obtained using smoothed empirical uncertainty distribution and p equations involving the parameters are constructed with the help of true uncertainty distribution function. Solving these parametric equations, the required estimates are found.

In this paper, the percentile matching method has been employed to estimate the parameters in linear uncertainty distribution. The underlying steps are explained below.

Let ξ be a linear uncertain variable $\xi \sim \mathcal{L}(a, b)$ with two unknown parameters a and b which are two positive real numbers satisfying $a < b$. Since there are two parameters in the linear uncertainty distribution function, the method of percentile matching uses two percentiles of predefined orders p_1 and p_2. On making use of the empirical uncertainty distribution, empirical percentiles of orders p_1 and p_2 are obtained and quantile values denoted by x_1 and x_2, respectively, are computed. It may be noted that $x_1 = \Phi_{n^{-1}}(p_1)$ and $x_2 = \Phi_{n^{-1}}(p_2)$. Two equations involving the parameters a and b are formed using x_1 and x_2 in the true uncertainty distribution function Φ. Solving for the parameters from the resulting equations, the percentile matching estimates of the parameters are obtained.

In the following section, a detailed study has been carried out on the estimation of parameters in linear uncertainty distribution. It gives a comparison of the performance of method of percentile matching and method of moments in estimation procedure.

Experimental Study

It is to be noted that while using percentile matching method, the estimated values and hence the accuracy of estimates depends on the orders of the percentiles used in estimation process. Hence, it is necessary to use percentiles of appropriate order to enhance the quality of estimates. In this section, it is proposed to make a detailed study on this aspect with reference to estimation of parameters in linear uncertainty distribution using experimental data sets. The main objective is to explore whether it is possible to identify optimal orders of percentiles which can be used for estimating the parameters appearing in linear uncertainty distribution based on numerical studies.

The error involved in the estimation of uncertain parameters can be measured by the quantity $\sum_{i=1}^{n}\left|\hat{\Phi}(x_i)-\alpha_i\right|$ where $\hat{\Phi}(x_i)'s$ are the estimated belief values and $\alpha_i's$ are the corresponding experimental belief values. In further discussion, it will be denoted by AE. That is,

$$\text{AE} = \sum_{i=1}^{n}\left|\hat{\Phi}(x_i)-\alpha_i\right|. \tag{5}$$

In the comparative study, experimental data sets simulated by a random mechanism are used. The process of generating one set of experimental data associated with the linear uncertainty distribution $\mathcal{L}(a, b)$ for a pre-fixed a and b is explained below.

(i) Determine a sequence of n equally spaced values in the interval (a,b), say $x_1, x_2, \ldots x_n$.

(ii) Compute the values of linear uncertainty distribution Φ $x_i's$ obtained in step (i) for $\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_n)$.

(iii) $\Phi(x_i)$ values obtained in step (ii) are either added with or subtracted in a randomized manner by $\varepsilon = \frac{\Phi(x_2)-\Phi(x_1)}{c}$ where c is a positive integer greater than 1. It is to be noted that $\Phi(x_i) - \Phi(x_{i-1})$ is a constant for all i since the distribution is linear. This leads to a sequence of values, namely, $\Phi(x_1) \pm \varepsilon, \Phi(x_2) \pm \varepsilon, \ldots, \Phi(x_n) \pm \varepsilon$.

These values are taken as belief degree values $\alpha_1, \alpha_2, \ldots, \alpha_n$ corresponding to $x_1, x_2, \ldots, x_n$ leading to the expert's experimental data set $(x_1, \alpha_1), (x_2, \alpha_2), \ldots, (x_n, \alpha_n)$.

In the experimental study, for a and b, 20 pairs of values with varying differences, namely, *10, 20, 30,* and *50* (each comprising five sets) as given in Table 1 are used. In order to reach reliable conclusions from the numerical study, for one set of values of parameters a and b, experimental data sets are generated by repeating the procedure mentioned above *100* times by using randomly generated ε's (described in step (iii) of the method of simulation explained above) taking the value of c as *3*. The values of parameters a and b are estimated for each set of expert's experimental data using method of moments and method of percentile matching for every choice of p_1 and p_2 which are determined by the procedure given below.

Table 1 Choices of intervals

Differences			(a,b)		
10	(5,15)	(10,20)	(15,25)	(20,30)	(25,35)
20	(5,25)	(10,30)	(15,35)	(20,40)	(25,45)
30	(10,40)	(15,45)	(20,50)	(25,55)	(30,60)
50	(10,60)	(15,65)	(20,70)	(25,75)	(30,80)

Values ranging from $\min(\alpha) + 0.01$ to 0.5 in step 0.01 are assigned for p_1 and for a given p_1, values ranging from $p_1 + 0.01$ to $\max(a) – 0.0001$ are assigned to p_2 in step *0.01*.

For each pair of p_1 and p_2, the value of AE is computed and the pair corresponding to minimum *AE* is recorded as the best pair of percentiles for the data being used. The *AE* due to the use of such best pair of percentiles is denoted by *AEP*. For each data, the method of moments as well as the method of least squares are also applied. The resulting Absolute Errors denoted by *AEM* and *AELS* are obtained.

The numerical study carried out by following the above procedure for one set of values of parameters is explained in detail below.

Consider the interval (a,b) of length 10 by taking $a = 5$ and $b = 15$ and assume x_i, $i = 1, 2, ..., 10$ takes values in *(5,15)*. The x values defined by $a + (b – a) \quad 0.10 \quad x$ are obtained as *6.00, 6.89, 7.78, 8.67, 9.56, 10.44, 11.33, 12.22, 13.11*, and *14.00* with uncertainty distribution function values (as provided by linear uncertainty distribution) *0.10, 0.19, 0.28, 0.37, 0.46, 0.54, 0.63, 0.72, 0.81*, and *0.90*. In this case, the ε value is found to be *0.03*. The belief degree values *0.07, 0.16, 0.31, 0.34, 0.43, 0.57, 0.66, 0.69, 0.78*, and *0.93* are obtained by randomly adding and subtracting the ε value with the linear uncertainty distribution function values. Thus, the experimental data set obtained is *(6.00,0.07)*, *(6.89,0.16)*, *(7.78,0.31)*, *(8.67,0.34)*, *(9.56,0.43)*, *(10.44,0.57)*, *(11.33,0.66)*, *(12.22,0.69)*, *(13.11,0.78)*, *(14.00,0.93)*. Using percentile matching method for estimation of parameters, the minimum absolute error *AEP* was found to be *0.24* for the above experimental data set considered with the orders of best choices of the percentiles p_1 and p_2 being *0.07* and *0.13*. The best percentile matching estimates of a and b are found to be *5.3* and *15.3*, respectively. The moment estimates of the parameters a and b are obtained as *5.37* and *14.74* with absolute error (*AEM*) *0.24*. The least square estimates of a and b are found to be *5.21* and *14.90* with *AELS 0.26*.

The process of finding the best pair, *AEP* and *AEM*, is repeated for all the *100* experimental data sets generated using the interval (a,b). It was found that in all simulated data sets, the value of *AEP* is less when compared to the values of *AEM* and *AELS*. Careful analysis over the best choice of percentiles did not lead to any conclusive evidence towards a universally best choice for each one of the intervals considered in the numerical study. One can think of different approaches for analyzing the results obtained in the numerical study in order to find the best pair of percentiles for one set of parametric values. It is reasonable to expect the level of deviation (created through ε) between the simulated belief levels, and the uncertainty distribution values have impact on the ultimate values of *AE*. That is, *AE* is likely to depend on the pattern followed in the simulation based on uncertainty distribution values. It is to be mentioned that the pattern followed in simulation can be quantified using entropy of the distribution of $+ \varepsilon$ and $– \varepsilon$. The entropy based on the distribution of $+ \varepsilon$ and $– \varepsilon$ is defined as

$$e = -\sum_{i=1}^{2} p_i \log_2(p_i) = -p_1 \log_2(p_1) - p_2 \log_2(p_2)$$

where p_1 and p_2 are the proportions of $+\varepsilon$ and $-\varepsilon$ generated while simulating belief degree values. Entropy value *0* indicates that the belief values are obtained by a complete shift of $+\varepsilon$ or $-\varepsilon$ from the uncertainty distribution function. On the other hand, the entropy value becomes higher if the number of positive and negative shifts tends to be equal.

In order to get an insight into results obtained in the experimental study, output related to *10* data sets simulated from (*5,15*) when *x* takes *10* values is provided in Table 2. Entries in a row are values of the absolute errors due to three different methods of estimation considered in this work along with the estimated values of parameters. It may be noted that the entropy value reported in a row is determined by using the distribution of $+\varepsilon$ and $-\varepsilon$ in the simulation process. From Table 2, it is clear that *AEP* due to method of percentile matching is always less than the moment error (*ME*) due to method of moments and least square error (*LSE*) due to method of least squares. Further, there are two cases where *AELS* happens to be equal to *AEP* due to error in approximation.

It is to be noted that the set of possible entropy values differ according to the number of values generated for experts' opinion. The best values of p_1 and p_2 are grouped according to the entropy values, and the weighted average of best percentile orders is computed for different values of entropies. The weighted average is considered since it is not necessary that the frequencies of occurrence of different entropy values differ. The frequency of occurrence of an entropy value in the simulated set is treated as the weight of that possible value. In this study, six different lengths which originate from various intervals of lengths *10*, *20*, *30*, and *50* are considered as values provided to the experts for eliciting their belief levels. Generally, the best choices of p_1 start from a smaller value (around *0.12*) and increases (up to value around *0.18*) as the entropy values increase up to a point and start decreasing (towards a value around *0.15*) beyond that value. In all the simulated data sets, it was observed that the optimal choices of p_2 exhibit an increasing pattern irrespective of the number of values generated from different intervals for expert's opinion. It increases from *0.30* to *0.70*. When the difference between the optimal orders as the entropy values change is apparent in the case of p_2,

Table 2 Sample output

Sl number	Entropy	p_1	p_2	$\hat{a}$	$\hat{b}$	AEP	AEM	AELS
1	0.97	0.07	0.13	5.3	15.3	0.23	0.24	0.26
2	0.97	0.31	0.66	4.38	15.42	0.21	0.3	0.24
3	0.97	0.07	0.84	5.35	14.58	0.12	0.14	0.12
4	0.97	0.16	0.84	5.43	14.57	0.22	0.23	0.24
5	0.97	0.25	0.72	5.3	15.3	0.23	0.26	0.27
6	0.97	0.13	0.42	4.59	15.49	0.21	0.32	0.23
7	1	0.5	0.78	4.48	15.53	0.13	0.29	0.13
8	0.88	0.39	0.92	4.7	14.7	0.17	0.22	0.21
9	0.88	0.07	0.36	5.3	15.3	0.17	0.26	0.24
10	0.88	0.13	0.7	4.57	15.61	0.15	0.27	0.15

such conclusion could not be arrived in the case of p_1. To illustrate this, box plots of the values of optimal orders corresponding to different entropy values for the cases of providing five and ten values for eliciting experts' opinion are presented in Figs. 1 and 2, respectively.

In order to examine the significance of differences between p_1 values with respect to variation in entropy values, one-way ANOVA was performed for each parametric setting of a and b. It may be noted that the cases where the entropy assumes the value *0* practically have no meaning, because no one will think of fitting a curve which completely lies either fully above or fully below the points in the experimental data. Hence, such values are excluded while performing analysis of variance. It was found that differences between the p_1 values are statistically significant with respect to the entropy values irrespective of the number of values provided to experts for expressing their belief levels under all choices of the parameters. Hence, Tukey Honest Significance Difference (HSD) test has been carried out to reach conclusive evidence. It was observed that the pairs (*0.72,0.97*), (*0.65,1*), (*0.59,0.98*), (*0.81,0.95*), and (*0.91,0.99*) are the pairs of entropy values in which optimal choices of p_1 happened to be equal for the cases of experts being provided with *5, 6, 7, 8*, and *9* values. In the case of *10* values, the two pairs (*0.72,1*) and (*0.88,0.97*) use equal values for p_1.

The entries in Table 3 can be used as guidance for deciding the appropriate order of percentiles to be used in the process of estimation of parameters in the case of linear uncertainty distribution.

Conclusion

In this paper, the utility of the method of percentile matching is investigated for estimating the parameters in linear uncertainty distribution. A detailed study on identifying optimal orders of percentiles to be used has been carried out numerically. Based on the experimental study, it is concluded that there is no globally optimum choices for the percentiles p_1 and p_2. The optimal choices of the percentiles depends on the number of values provided to the experts for obtaining their belief levels as well as the pattern present in the experimental data set. The patterns present in the data set are gauged with the help of entropy values as explained earlier. The entries in Table 3 can be used for deciding suitable orders of percentiles.

Even though the study is confined to linear uncertainty distribution, it can be extended in similar fashion to other uncertainty distributions as well by using appropriate number of percentiles. The superiority of the percentile method over the method of

Fig. 1 Box plot of distribution of percentiles for data sets with *5* values
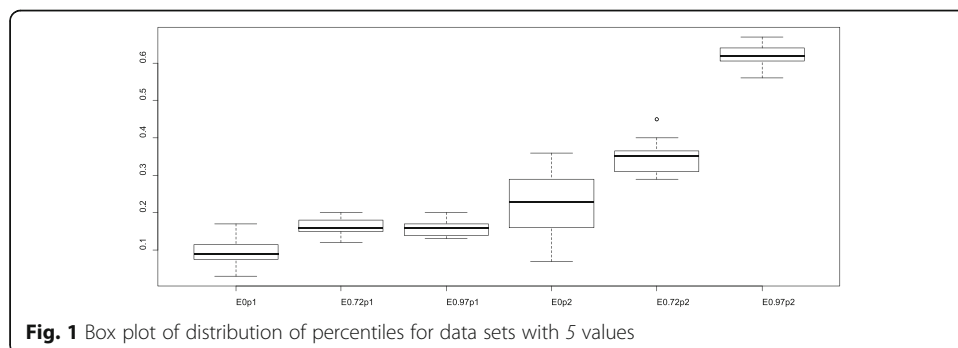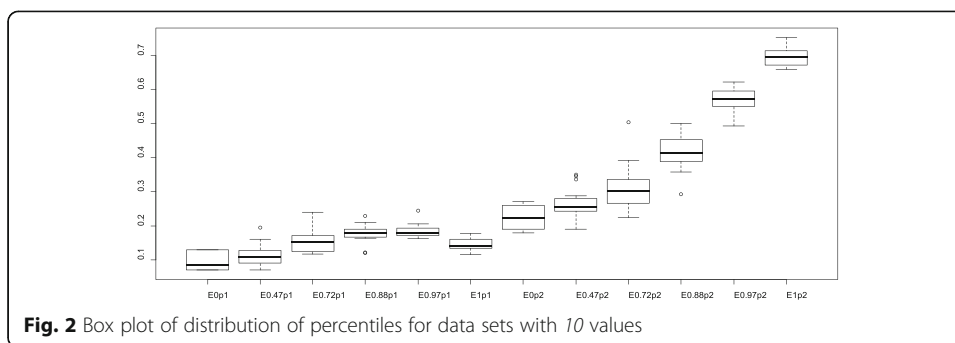
Fig. 2 Box plot of distribution of percentiles for data sets with *10* values

Table 3 Recommended values

Number of values							
		Entropy	0.72	0.97			
	5	p_1	0.16	0.16			
		p_2	0.35	0.62			
		Entropy	0.65	0.92	1		
	6	p_1	0.17	0.18	0.16		
		p_2	0.31	0.52	0.72		
		Entropy	0.59	0.86	0.99		
	7	p_1	0.16	0.18	0.16		
		p_2	0.33	0.45	0.66		
		Entropy	0.54	0.81	0.95	1	
	8	p_1	0.15	0.18	0.18	0.14	
		p_2	0.31	0.39	0.58	0.7	
		Entropy	0.5	0.76	0.92	0.99	
	9	p_1	0.13	0.16	0.18	0.18	
		p_2	0.28	0.34	0.46	0.66	
		Entropy	0.47	0.72	0.88	0.97	1
	10	p_1	0.12	0.16	0.18	0.18	0.15
		p_2	0.27	0.31	0.41	0.57	0.7

moments and the method of least squares has been established through extensive numerical study. In the present study, we have used a kind of search procedure to determine the optimal orders of the percentiles being used. One can use soft computing algorithms like genetic algorithm, ant colony optimization to identify the optimal choices, and correct form of uncertainty distribution. The authors are working in that direction.

Authors' contributions
SS conceived the idea of using percentile matching, designed the framework and overall organization of the paper. KA helped in the preparation of the manuscript and conducting simulation studies. Both authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Klugman, SA, Panjer, HH, Wilmot, GE: Loss models: from data to decisions, 3rd ed, John Wiley & Sons, Ontario (2008)
2. Liu, B.: Uncertainty theory, 2nd edn. Springer, Berlin (2007)

3.   Liu, B.: Uncertainty theory: a branch of mathematics for modeling human uncertainty. Springer, Berlin (2010)
4.   Liu, B: Why is there a need for uncertainty theory? J. Uncertain Syst. **6**(1), 3-10 (2012)
5.   Liu, B: Uncertainty theory, 5th ed, http://orsc.edu.cn/liu/ut.pdf.
6.   Sheng, Y.H., Kar, S.: Some results of moments of uncertain variable through inverse uncertainty distribution. Fuzzy Optim. Decis. Making **14**, 57–76 (2015)
7.   Wang, XS, Peng ZX: Method of moments for estimating uncertainty distributions. J. Uncertainty Analys. Appl. **2**(5), (2014)
8.   Zhang, Z.M.: Some discussions on uncertain measure. Fuzzy Optim. Decis. Making **10**(1), 31–43 (2011)