References
References


Published Papers
Percentile Matching Estimation of Uncertainty Distribution

S Sampath* and K. Anjana

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**


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

\[
\mathcal{M} \left\{ \prod_{k=1}^{\infty} \Lambda_k \right\} = \prod_{k=1}^{\infty} \mathcal{M}_k \{ \Lambda_k \}
\]

where \(\Lambda_k\) 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 \(\xi\) 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 \(\Phi\) of an uncertain variable \(\xi\) is defined by Liu [2] as

\[
\Phi(x) = \mathcal{M}\{\xi \leq x\}, \forall x \in \mathbb{R}.
\]

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 \(\xi\) is called linear if it has uncertainty distribution of the form

\[
\Phi(x) = \begin{cases} 
0, & \text{if } x \leq a \\
\frac{x-a}{b-a}, & \text{if } a \leq x \leq b \\
1, & \text{if } x \geq b
\end{cases}
\]

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

\[
\Phi_n(x) = \begin{cases} 
0, & \text{if } x < x_1 \\
\frac{a_i + (a_{i+1}-a_i)(x-x_i)}{x_{i+1}-x_i}, & \text{if } x_i \leq x \leq x_{i+1}, 1 \leq i < n \\
1, & \text{if } x > x_n
\end{cases}
\]

where \(x_1 < x_2 < \ldots < x_n\) and \(0 \leq a_1 \leq a_2 \leq \ldots \leq a_n \leq 1\).

**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 \(\xi\) is defined by Liu [2] as

\[
E[\xi] = \int_{-\infty}^{+\infty} \mathcal{M}\{\xi \geq x\} \, dx - \int_{-\infty}^{0} \mathcal{M}\{\xi \leq x\} \, dx
\]

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

$$E[\xi] = \int_0^1 \Phi^{-1}(a) \, da.$$  

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

Let $\xi$ be an uncertain variable with uncertainty distribution $\Phi$. Then by Liu [5],

(i) If $k$ is an odd number, then the $k$th moment of $\xi$ is defined as

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

(ii) If $k$ is an even number, then the $k$th moment of $\xi$ is defined as

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

(iii) If $k$ is a positive integer, then the $k$th moment of $\xi$ is defined as

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

Sheng and Kar [6] proved that, if an uncertain variable $\xi$ has a regular uncertainty distribution $\Phi$ and $k$ is a positive integer, then the $k$th moment of $\xi$ is

$$E[\xi^k] = \int_0^1 (\Phi^{-1}(a))^k \, da.$$  

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.

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

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

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

For a given expert’s experimental data

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

that meet the condition

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

where $x_r$’s are the observed values and $\alpha_r$’s are the respective belief degree values, Wang and Peng [7] gives the $k$th empirical moment of the uncertain variable $\xi$ based on empirical uncertainty distribution as
\[ F^k = a_1 x_1^k + \frac{1}{k+1} \sum_{i=1}^{n-1} \sum_{j=0}^{k} (a_{i+1} - a_i) x_i^{k-j} x_{i+1}^j + (1-a_n)x_n^k. \] (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 \( k \)th percentile is found by interpolating between the two data points that are around the \( n \)th observation. A percentile matching estimate of the vector valued parameter \( \theta = (\theta_1, \theta_2, \ldots, \theta_p) \) is any solution of the \( p \) equations \( \pi_{g_k}(\theta) = \hat{\pi}_{g_k}; k = 1, 2, 3, \ldots, p \) where \( g_1, g_2, \ldots, g_p \) are \( p \) arbitrarily chosen percentiles. Here, the left hand side, namely, \( \pi_{g_k}(\theta) \), represents the theoretical percentile of order \( g_k (k = 1, 2, 3, \ldots, p) \) 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 \( h = (n+1)g - j \). Here, \( \lfloor \cdot \rfloor \) indicates the greatest integer function and \( x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)} \) are the order statistics from the sample.

It is to be noted that \( F(\pi_{g_k} \mid \theta) = g_k; k = 1, 2, 3, \ldots, p \) where \( F(.) \) is the true cumulative distribution of the underlying probability distribution. On substituting \( \hat{\pi}_{g_k} \) in lieu of \( \pi_{g_k} \) 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

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

with respect to \( \theta \).

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

Let a non-negative uncertain variable \( \xi \) 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 \( 0 \leq x_1 < x_2 < \ldots < x_n, 0 \leq \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n \leq 1 \).

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

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

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

For example, let \( \xi \) 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
\[ \Phi(x) = \begin{cases} 
0, & \text{if } x \leq a \\
\frac{x-a}{b-a}, & \text{if } a \leq x \leq b \\
1, & \text{if } x \geq b. 
\end{cases} \]

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 \( \xi_1 \) and \( \xi_2 \) 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

\[(x_1, a_1), (x_2, a_2), \ldots, (x_n, a_n)\]

where \( x_i, i = 1, 2, 3, \ldots, n \) are the observed values and \( a_i, i = 1, 2, 3, \ldots, n \) are the belief degrees. Here, it is assumed that \( 0 \leq x_1 < x_2 < \ldots < x_n \) and \( 0 \leq a_1 \leq a_2 \leq \ldots \leq a_n \leq 1 \).

The observed values \( x_i, i = 1, 2, 3, \ldots, n \) 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 \( \xi \) 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 \( \Phi \). 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 

\[ \sum_{i=1}^{n} |\hat{\Phi}(x_i) - \alpha_i| \]

where \(\hat{\Phi}(x_i)\) 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,

\[ AE = \sum_{i=1}^{n} |\hat{\Phi}(x_i) - \alpha_i| . \]  

(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 \(\Phi\) \(x_i\)'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 \(\varepsilon = \frac{\Phi(x_i) - \Phi(x_{i-1})}{c}\) 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 \(\varepsilon\)'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.
Values ranging from $\min(a) + 0.01$ 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 $a + (b - a) \times 0.10 \times x$ 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 $\varepsilon$ 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 $\varepsilon$ 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 $\varepsilon$) 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

**Table 1** Choices of intervals

<table>
<thead>
<tr>
<th>Differences</th>
<th>$(a,b)$</th>
<th>$(a,b)$</th>
<th>$(a,b)$</th>
<th>$(a,b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>(5,15)</td>
<td>(10,20)</td>
<td>(15,25)</td>
<td>(20,30)</td>
</tr>
<tr>
<td>20</td>
<td>(5,25)</td>
<td>(10,30)</td>
<td>(15,35)</td>
<td>(20,40)</td>
</tr>
<tr>
<td>30</td>
<td>(10,40)</td>
<td>(15,45)</td>
<td>(20,50)</td>
<td>(25,55)</td>
</tr>
<tr>
<td>50</td>
<td>(10,60)</td>
<td>(15,65)</td>
<td>(20,70)</td>
<td>(25,75)</td>
</tr>
</tbody>
</table>

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 $\varepsilon$) 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

**Table 1** Choices of intervals

<table>
<thead>
<tr>
<th>Differences</th>
<th>$(a,b)$</th>
<th>$(a,b)$</th>
<th>$(a,b)$</th>
<th>$(a,b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>(5,15)</td>
<td>(10,20)</td>
<td>(15,25)</td>
<td>(20,30)</td>
</tr>
<tr>
<td>20</td>
<td>(5,25)</td>
<td>(10,30)</td>
<td>(15,35)</td>
<td>(20,40)</td>
</tr>
<tr>
<td>30</td>
<td>(10,40)</td>
<td>(15,45)</td>
<td>(20,50)</td>
<td>(25,55)</td>
</tr>
<tr>
<td>50</td>
<td>(10,60)</td>
<td>(15,65)</td>
<td>(20,70)</td>
<td>(25,75)</td>
</tr>
</tbody>
</table>
\[ e = -\sum_{i=1}^{2} p_i \log_2(p_i) = -p_1 \log_2(p_1) - p_2 \log_2(p_2) \]

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

Received: 19 May 2016 Accepted: 30 September 2016
Published online: 17 October 2016

References

4. Liu, B.: Why is there a need for uncertainty theory? J. Uncertain Syst. 6(1), 3-10 (2012)


Percentile Matching Estimation of Zigzag Uncertainty Distribution

S. Sampath, Department of Statistics, University of Madras, Chennai, India
K. Anjana, University of Madras, Chennai, India

ABSTRACT

The problem of estimating parameters involved in zigzag uncertainty distribution is considered in this article. Sensing the difficulties involved in the direct application of statistical estimation techniques for uncertainty distribution, the present article considers the application of the method of percentile matching for estimating the unknown parameters of zigzag uncertainty distribution. This article clearly establishes the fact that the percentile matching method gives better estimates when compared to the method of moments if sample percentiles of appropriate orders are used in the estimation process. Detailed numerical studies have been carried out using simulated datasets possessing different characteristics for identifying optimal orders of percentiles which give better estimates of parameters.

KEYWORDS

Method of Moments, Method of Percentile Matching, Zigzag Uncertainty Distribution

1. INTRODUCTION

Different types of uncertainties arise in real life situations. According to Liu (2008) randomness and impreciseness (fuzziness) are basic types of objective uncertainty and subjective uncertainty, respectively. Probability theory has been developed to handle random phenomena in which the events are well defined and considered not to have vagueness or uncertainty. The idea of fuzzy set theory has been introduced by Zadeh (1965) in order to deal with fuzziness through membership values. Later the concept of fuzzy graphs has been introduced by Rosenfield (1975). Fuzzy robust graph coloring problem has been discussed in Dey, Pradhan, Pal and Pal (2015). Vertex coloring of a fuzzy graph using alpha cut can be seen in Dey and Pal (2012). Interval type 2 fuzzy set in fuzzy shortest path problem is available in Dey, Pal and Pal (2016). Several works on the applications of fuzzy set theory have been carried out in different branches of statistics.

Liu (2007) introduced the concept of Uncertainty theory. According to Liu (2017), the concept of uncertainty theory is one of the options available to deal with indeterminate phenomena whose outcomes cannot be predicted in advance. It turned out to be the solution for problems in the contexts where no samples are available which creates difficulty in using probability theory for dealing with such situations. In such cases, opinions of the domain experts become the only choice for further study. Liu (2017) framed uncertainty theory to model the belief degrees of domain experts in various contexts. Belief degree refers to the belief levels of experts regarding the occurrence of particular events. Structural characteristics of uncertain measure have been discussed in Zhang (2011). One
can find similarities in the process of developing uncertainty theory with that of probability theory. However, the ideas developed in uncertainty theory find applications in dealing with problems arising out of impreciseness created in non-stochastic manner.

Liu (2017) pioneered the uncertainty theory over the years and has developed several study areas similar to that existing in probability theory. Uncertain Measure, Uncertain Variable, Uncertain Programming, Uncertain Risk Analysis, Uncertain Reliability Analysis, Uncertain Propositional Logic, Uncertain Set, Uncertain Logic, Uncertain Inference, Uncertain Process, Uncertain Calculus, Uncertain Differential Equation, Uncertain Finance and Uncertain Statistics are some of the concepts developed under uncertainty theory. Liu (2017) gives a detailed explanation of these concepts. Studies on testing uncertain hypotheses about uncertainty distribution functions have been made by Wang, Gao and Guo (2012) and Sampath and Ramya (2013). The concept of uncertain random variables has been introduced by Liu (2013) as a mixture of uncertainty and randomness.

The estimation of parameters in uncertainty distributions has received the attention of researchers. Liu (2017) considered the least square estimates of parameters in linear uncertainty distribution for a given expert’s experimental data found with the help of MATLAB toolbox. Method of least squares, Method of moments and Delphi method are some of the existing methods of estimation available in the uncertainty literature. Wang and Peng (2014) explains the method of moments and gives an expression for finding the empirical moments of uncertain variables. Sheng and Kar (2015) derived the first two theoretical moments of a linear uncertain variable using the idea of inverse uncertainty distribution and also explains the central moments of an uncertain variable. One of the important methods for estimation available in the theory of estimation namely, method of maximum likelihood cannot be used due to the absence of the concept of density functions in uncertainty theory. Hence, alternative approaches for estimation of parameters involved in uncertainty distributions become necessary. Motivated by the method of percentile matching available in statistical theory of estimation, recently, Sampath and Anjana (2016) introduced the method of percentile matching for estimating the unknown parameters in Linear uncertainty distribution. The details of method of percentile matching used in statistical theory of estimation can be seen in Klugman, Panjer and Wilmot (2008). Following the approach pursued in the work related to the estimation of unknown parameters in linear uncertainty distribution, this article is devoted for studying the problem related to the estimation of unknown parameters involved in Zigzag uncertainty distribution. This article is organized as follows. Second section gives a detailed explanation on various basic concepts in uncertainty theory. Third section explains the method of percentile matching in detail. Fourth section describes different methods of estimation available in uncertainty theory. Detailed description of the experimental study carried out for the estimation of unknown parameters in zigzag uncertainty distribution is given in the fifth section. Findings and conclusions are presented in the sixth section.

2. UNCERTAINTY THEORY

This section discusses about various basic concepts in uncertainty theory with reference to Liu (2017).

Consider \( \Gamma \) as a nonempty set and \( \mathcal{L} \) be a \( \sigma \)-algebra over \( \Gamma \). Each element \( \Lambda \in \mathcal{L} \) is called an event. A number \( \mathcal{M}(\Lambda) \) indicates the level of occurrence of the event \( \Lambda \).

2.1. Uncertain Measure

A set function \( \mathcal{M} \) is said 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:
\( M \left( \bigcup_{i=1}^{\infty} \Lambda_i \right) \leq \sum_{i=1}^{\infty} M \{ \Lambda_i \} \)

Even though probability measure satisfies the first three axioms, 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 \( k = 1, 2, 3, \ldots \). Then the product uncertain measure is an uncertain measure satisfying the condition:

\[
M \left( \prod_{k=1}^{\infty} \Lambda_k \right) = \bigwedge_{k=1}^{\infty} M \{ \Lambda_k \}
\]

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

### 2.2. Uncertain Variable

A measurable function \( \xi \) from an uncertainty space \( (\Gamma, \mathcal{L}, \mathcal{M}) \) to the set of real numbers is said to be an uncertain variable if \( \{ \xi \in B \} \) is an event for any Borel set \( B \) of real numbers.

### 2.3. Uncertainty Distribution

Uncertainty distribution \( \Phi \) of an uncertain variable \( \xi \) is defined as:

\[
\Phi(x) = M \{ \xi \leq x \}, \quad \forall x \in \mathbb{R} \tag{1}
\]

Various uncertainty distributions namely, Linear, Zigzag, Normal and Lognormal are available in uncertainty literature. This work emphasizes the study on zigzag uncertainty distribution.

### 2.4. Zigzag Uncertainty Distribution

An uncertain variable \( \xi \) is called Zigzag if it has a zigzag uncertainty distribution:

\[
\Phi(x) = \begin{cases} 
0, & \text{if } x \leq a \\
\frac{x-a}{2(b-a)}, & \text{if } a \leq x \leq b \\
\frac{x+2b-c}{2(c-a)}, & \text{if } b \leq x \leq c \\
1, & \text{if } x \geq c 
\end{cases} \tag{2}
\]

denoted by \( Z\{a, b, c\} \) where \( a, b, c \) are real numbers with \( a < b < c \).

Empirical uncertainty distribution function plays a key role in the estimation process of uncertainty distributions.

Let \( (x_1, \alpha_1), (x_2, \alpha_2), \ldots, (x_n, \alpha_n) \) be a given expert’s experimental data provided:
where $x_i$'s are the observed values and $\alpha_j$'s being the expert's belief degree levels. In this work, we shall refer the number of values presented to an expert, namely $n$ as response cardinality.

### 2.5. Empirical Uncertainty Distribution

Empirical uncertainty distribution for a given experimental data is defined as:

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

### 2.6. Regular Uncertainty Distribution

An uncertainty distribution $\Phi(x)$ is called regular if it is a continuous and strictly increasing function with respect to $x$ where $0 < \Phi(x) < 1$ and $\lim_{x \to -\infty} \Phi(x) = 0$, $\lim_{x \to +\infty} \Phi(x) = 1$. Linear, Zigzag, Normal, Lognormal uncertainty distributions are all examples of regular uncertainty distributions.

### 2.7. Expected Value of an Uncertain Variable

The expected value of an uncertain variable $\xi$ is defined as:

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

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

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

From the above expression, using integration by parts:

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

If $\xi$ has a regular uncertainty distribution $\Phi$ then by substituting $\Phi(x)$ with $\alpha$, $x$ with $\Phi^{-1}(\alpha)$ in the previous expression and following the change of variables of integral:

$$
E[\xi] = \int_{0}^{1} \Phi^{-1}(\alpha) \, d\alpha
$$
2.8. Moments (Liu, 2007)

If $\xi$ is an uncertain variable and $k$ is a positive integer, then the $k^{th}$ moment of $\xi$ is defined as $E \left[ \xi^k \right]$.

Let $\xi$ be an uncertain variable with uncertainty distribution $\Phi$. Then, from Liu (2017):

(i) If $k$ is an odd number, then the $k^{th}$ moment of $\xi$ is defined as:

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

(ii) If $k$ is an even number, then the $k^{th}$ moment of $\xi$ is defined as:

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

(iii) If $k$ is a positive integer, then the $k^{th}$ moment of $\xi$ is defined as:

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

If an uncertain variable $\xi$ has a regular uncertainty distribution $\Phi$ and $k$ is a positive integer, then Sheng and Kar (2015) defines the $k^{th}$ moment of $\xi$ as:

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

First three theoretical moments of zigzag uncertain variable $\xi$ derived by Liu (2017) are listed below:

$$E [\xi] = \frac{a + 2b + c}{4} \quad (4)$$

$$E [\xi^2] = \frac{a^2 + ab + 2b^2 + bc + c^2}{6} \quad (5)$$

$$E [\xi^3] = \frac{a^3 + a^2b + ab^2 + 2b^3 + b^2c + bc^2 + c^3}{8} \quad (6)$$

Based on the experimental data one can develop empirical moments of different orders. The $k^{th}$ empirical moment of the uncertain variable $\xi$ based on empirical uncertainty distribution as given by Wang and Peng (2014) is:
\[
\xi^k = \alpha_1 x_1^k + \frac{1}{k+1} \sum_{i=1}^{n-1} \sum_{j=0}^{k} (\alpha_i - \alpha_i) x_i^j x_{i+1}^{k-j} + (1 - \alpha_n) x_n^k
\]

(7)

Thus, we have presented the basic terminology as well as some useful results available in Uncertainty theory. The following section explains the method of percentile matching which is primarily considered in this work.

3. METHODS FOR ESTIMATING THE PARAMETERS IN PROBABILITY DISTRIBUTIONS

Several methods are available in the statistical theory of estimation for estimating the unknown parameters in probability distributions. It includes the method of least squares, method of moments, method of maximum likelihood and method of percentile matching. Method of least squares minimizes the squared distance between the observed data and their fitted values for estimating the parameters. Method of moments is one of the oldest methods available in the statistical theory of estimation. It gives estimates of parameters on equating the first \( k \) population and the corresponding sample moments and solving the resulting equations in terms of the sample moments. Method of percentile matching makes use of percentiles of different orders of available data for estimation of parameters in a distribution.

3.1. Method of Percentile Matching

In case of discrete sample data, percentiles are found by smoothing the data. If there are \( n \) observations then the \( k^{th} \) percentile is found by interpolating between the two data points that are around the \( nk^{th} \) observation. A percentile matching estimate of the vector valued parameter \( \theta = (\theta_1, \theta_2, ..., \theta_p) \) is any solution of the \( p \) equations \( \pi_{g_k}(\theta) = \pi_{g_k}, k = 1, 2, 3, ..., p \) where \( g_1, g_2, ..., g_p \) are \( p \) arbitrarily chosen percentiles. 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 \( \pi_{g_k} \) gives the corresponding smoothed empirical estimate. The expression \( \pi_{g_k} = (1 - h)x_{(j)} + hx_{(j+1)} \) is used to find the smoothed empirical estimate of a percentile where \( j = \lfloor (n+1)g \rfloor \) and \( h = (n+1)g - j \). \( \lfloor \cdot \rfloor \) indicates the greatest integer function. \( x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)} \) are the order statistics from the sample.

It is to be noted that:

\[
F\left(\pi_{g_k} | \theta \right) = g_k, k = 1, 2, 3, ..., p
\]

(8)

where \( F(.) \) is the true cumulative distribution of the underlying probability distribution. Equation (8) leads to a system of \( p \) equations formed by substituting \( \pi_{g_k} \) in place of \( \pi_{g_k} \). Solutions based on this system of equations are the estimates of the parameters. Here, \( \pi_{g_k} \) cannot be obtained for \( g < \frac{1}{n+1} \) or \( g > \frac{n}{n+1} \). This is reasonable as it is meaningless to find the value of very large or small percentiles from small samples. Smoothed version is used whenever an empirical percentile estimate is needed.
Next section gives a detailed explanation of the method of moments and method of percentile matching in estimating the unknown parameters of uncertainty distributions.

4. METHODS FOR ESTIMATING UNCERTAINTY DISTRIBUTIONS

4.1. Method of Moments

This method of estimating unknown parameters of an uncertainty distribution was proposed by Wang and Peng (2014). The method is explained below.

Let $\xi$ be a non-negative uncertain variable having an uncertainty distribution $\Phi(x|\theta_1, \theta_2, ..., \theta_p)$ with unknown parameters $(\theta_1, \theta_2, ..., \theta_p)$. Expression for the $k^{th}$ empirical moment of $\xi$ for a set of experimental data:

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

satisfying the conditions:

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

is given in (7). The first $p$ theoretical moments of $\xi$ are in general functions of the parameters involved in the uncertainty distribution. For example, as one can see from equations (4), (5) and (6), the theoretical moments of zigzag uncertainty distribution depend on the three parameters $a, b$ and $c$. These theoretical moments are equated to the corresponding empirical moments and the moment estimates $(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p)$ are found. It is equivalent to solving the system of equations:

$$\int_0^\infty \left(1 - \Phi \left(\frac{x}{\xi} \bigg| \theta_1, \theta_2, ..., \theta_p \right) \right) dx = \xi^k, k = 1, 2, 3, ..., p$$

where $\xi^1, \xi^2, ..., \xi^p$ are the empirical moments found using Equation (7).

Let $\xi$ be a zigzag uncertain variable following zigzag uncertainty distribution with three unknown parameters $a, b$ and $c$ which are real numbers such that $a < b < c$. Estimation of the three unknown parameters using method of moments is carried out using the first three theoretical and corresponding empirical moments of zigzag uncertain variable. The first three theoretical moments of zigzag uncertain variable are given by (4), (5) and (6). The corresponding empirical moments are found by substituting the values 1, 2 and 3 in (7). Then the first three theoretical and corresponding empirical moments are equated and the resulting three nonlinear equations are solved using the package *nleqslv* available in R. Solving the nonlinear equations gives the moment estimates of the unknown parameters $a, b$ and $c$.

4.2. Method of Percentile Matching

The method of percentile matching for estimating the unknown parameters in zigzag uncertainty distribution is discussed in detail below.

Consider a zigzag uncertain variable $\xi \sim \mathcal{Z}(a, b, c)$ where $a, b$ and $c$ are the three unknown parameters which are positive real numbers such that $a < b < c$. As there are three unknown
parameters in zigzag uncertainty distribution, this method makes use of three percentiles of predefined orders \( p_1, p_2 \) and \( p_3 \). Three empirical percentiles of orders \( p_1, p_2 \) and \( p_3 \) are obtained using empirical uncertainty distribution defined in (3). That is, the empirical percentiles \( x_1 = \Phi^{-1}(p_1) \), \( x_2 = \Phi^{-1}(p_2) \) and \( x_3 = \Phi^{-1}(p_3) \) are obtained. Substitution of the values of \( x_1, x_2 \) and \( x_3 \) in the true distribution function \( \Phi \) (given in (2)) gives three functional expressions involving the three parameters \( a, b \) and \( c \), namely, \( \Phi(x_1), \Phi(x_2) \) and \( \Phi(x_3) \). Equating them with \( p_1, p_2 \) and \( p_3 \) a system of three equations involving \( a, b \) and \( c \) namely, \( \Phi(x_1) = p_1, \Phi(x_2) = p_2 \) and \( \Phi(x_3) = p_3 \) are obtained. Solving these three equations, the percentile matching estimates are obtained.

The following section gives a detailed study of estimation of the zigzag uncertain parameters \( a, b \) and \( c \).

**5. EXPERIMENTAL STUDY**

In this section, a simulation based experimental study is carried out to assess the efficiency of the percentile matching method in estimating the three unknown parameters in zigzag uncertainty distribution. The estimation of parameters by method of percentile matching is carried out using three percentiles of predetermined order. The absolute error involved in the estimation of parameters is defined as:

\[
AE = \sum_{i=1}^{n} |\Phi(x_i) - \alpha_i|
\]

(9)

where \( \Phi(x_i) \)'s are the estimated belief values and \( \alpha_i \)'s are the corresponding true belief levels.

The efficiency of an estimation procedure depends on the orders of percentiles used in the estimation process. The percentiles which give minimum absolute error can be taken as optimal ones for estimating the unknown parameters \( a, b \) and \( c \). The technique of simulating an experimental data set and the procedure used for determination of optimal orders are explained below.

**5.1. Data Simulation**

The steps involved in simulating one set of expert’s experimental data set with response cardinality \( n \) from zigzag uncertainty distribution \( \mathcal{Z}(a,b,c) \) for a given \( a, b \) and \( c \) in a random manner is explained below.

Construct a sequence of \( n \) equally spaced values \( x_1, x_2, \ldots, x_n \) in the interval \( (a,c) \). For each, \( x_i, i = 1, 2, 3, \ldots, n \), compute the value of the zigzag uncertainty distribution denoted by \( \Phi_1, \Phi_2, \ldots, \Phi_n \). A quantity:

\[
\varepsilon = \min \left( \frac{\Phi(x_2) - \Phi(x_1)}{m}, \frac{\Phi(x_n) - \Phi(x_{n-1})}{m} \right)
\]

is either added with or subtracted from the zigzag uncertainty distribution values in a random manner. The value of \( \varepsilon \) is chosen in such a manner which ensures the distribution of the values in the interval \( (a,c) \). Here, \( m \) is any positive integer greater than 1. Thus, the sequence of values \( \Phi(x_1) \pm \varepsilon, \Phi(x_2) \pm \varepsilon, \ldots, \Phi(x_n) \pm \varepsilon \) generated are considered as the belief levels \( \alpha_1, \alpha_2, \ldots, \alpha_n \)
corresponding to the values \( x_1, x_2, \ldots, x_n \). This leads to one set of experimental data
\[
\left\{ x_1, \alpha_1 \right\}, \left\{ x_2, \alpha_2 \right\}, \ldots, \left\{ x_n, \alpha_n \right\}.
\]

To understand the simulation process clearly, details of the numerical study carried out for the case \( a = 5, b = 10 \) and \( c = 20 \) of length 15 is explained below. Note that, the parameter \( b \) lies closer to \( a \). A sequence of 5 equally spaced values 6.5, 9.5, 12.5, 15.5 and 18.5 ranging from 5 to 20 are determined whose zigzag uncertainty distribution values are 0.150, 0.450, 0.625, 0.775 and 0.925. Assuming the value of \( \varepsilon \) as 0.05, the simulated belief degree values are obtained by randomly adding and subtracting \( \varepsilon \) from the zigzag uncertainty distribution values. In our random simulation the values are obtained as 0.100, 0.500, 0.575, 0.725 and 0.975. Hence, the simulated experimental data set thus obtained is
\[
\{6.5, 0.100\}, \{9.5, 0.500\}, \{12.5, 0.575\}, \{15.5, 0.725\} \text{ and } \{18.5, 0.975\}.
\]

5.2. Optimal Orders Determination

The process of determination of optimal orders of percentiles to be used in the estimation of \( a, b \) and \( c \) is carried out using a search process by considering all possible \( p_1, p_2 \) and \( p_3 \) from a given range of values. The values of \( p_1 \) range from minimum value of alpha to 0.48 increased by step 0.01, \( p_2 \) from \( p_1 + 0.01 \) to maximum of alpha \(-0.0001 - 0.01 \) increased by step 0.01 and \( p_3 \) from \( p_2 + 0.01 \) to maximum of alpha \(-0.0001 \) increased by step 0.01. The value of \( AE \) as defined in (9) is computed for every choice of \( p_1, p_2 \) and \( p_3 \). The set of \( p_1, p_2 \) and \( p_3 \) corresponding to the minimum AE is chosen as optimal orders of percentiles.

The minimum absolute error due to method of percentile matching for the experimental data set is found to be 0.175. The best choice of orders of percentiles \( p_1, p_2 \) and \( p_3 \) are found to be 0.3, 0.5 and 0.89 respectively. The best percentile matching estimates of \( a, b \) and \( c \) are 5.75, 9.5 and 19.73.

In the experimental study, 18 intervals out of which 9 intervals of lengths 15, 25 and 35 are chosen in such a way that the unknown parameter \( b \) lies closer to \( a \) and the remaining 9 intervals of lengths 15, 25 and 35 are chosen such that \( b \) lies closer to \( c \). Intervals used in the study are listed in Table 1.

In order to ensure the reliability of the results obtained from the numerical study, for a given set of values of \( a, b \) and \( c \), 100 data sets are generated by repeating the above described simulation procedure. Here, the value of \( m \) is taken as 3. This exercise is repeated for each data set in all intervals.

The \( AEP \) and \( AEM \) obtained in all the 100 experimental datasets generated from each of these 18 intervals have been recorded. The abbreviation \( AEP \) is used to denote the \( AE \) corresponding to the best choice of \( p_1, p_2 \) and \( p_3 \). Method of moments is also applied for the experimental dataset and the \( AE \) due to this method denoted by \( AEM \) is computed. It was found that \( AEP \) is lesser than \( AEM \) in all cases. This observation holds good in all the 18 cases irrespective of the position of parameter \( b \) with respect to \( a \) and \( c \). Therefore, it is concluded that the method of percentile matching gives better estimates of \( a, b \) and \( c \) than those obtained in the method of moments irrespective of the distance of \( b \) from \( a \) and \( c \).

The choice of best orders of percentiles depends on the absolute error which in turn depends on the deviation of simulated belief levels from the uncertainty distribution values through \( \varepsilon \). Hence, it is decided to make use of the concept of entropy to quantify the level of this deviation by analysing the distribution of \( +\varepsilon \) and \( -\varepsilon \). The entropy for the distribution of \( +\varepsilon \) and \( -\varepsilon \) is defined as:
Table 1. Choices of intervals

<table>
<thead>
<tr>
<th>Length of the Intervals</th>
<th>Position of b</th>
<th>((a, b, c))</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>(b) closer to (a)</td>
<td>(5, 10, 20)</td>
</tr>
<tr>
<td></td>
<td>(b) closer to (c)</td>
<td>(5, 15, 20)</td>
</tr>
<tr>
<td>25</td>
<td>(b) closer to (a)</td>
<td>(5, 15, 30)</td>
</tr>
<tr>
<td></td>
<td>(b) closer to (c)</td>
<td>(5, 20, 30)</td>
</tr>
<tr>
<td>35</td>
<td>(b) closer to (a)</td>
<td>(5, 20, 40)</td>
</tr>
<tr>
<td></td>
<td>(b) closer to (c)</td>
<td>(5, 25, 40)</td>
</tr>
</tbody>
</table>

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

Here, \(p_1\) and \(p_2\) denote the proportions of the cases where \(\varepsilon\) gets added and subtracted respectively from the true zigzag uncertainty distribution. The entropy value 0 indicates a complete shift of \(+\varepsilon\) or \(-\varepsilon\) from the uncertainty distribution function and higher value indicates equal number of positive and negative shifts.

To get an understanding about the nature of results obtained, outcomes of experimental study related to 10 datasets generated from interval \((5, 10, 20)\) and \((5, 15, 20)\) with response cardinality 5 are given in Table 2 and Table 3 respectively.

In Tables 2 and 3, the \(AEP\) columns give the minimum of the absolute errors when percentile matching method is used and the \(AEM\) columns provide the absolute error corresponding to the method of moments. Further, the orders of percentiles corresponding the minimum \(AEP\) namely, \(p_1, p_2\) and \(p_3\) as well as the estimated values \(\hat{a}, \hat{b}\) and \(\hat{c}\) are also presented. The whole procedure is repeated for the optimal orders of percentiles in all 18 intervals having response cardinality ranging from 5 to 9.

In order to understand the behaviour of the distribution of optimal orders of percentiles in all intervals, boxplots have been drawn for response cardinalities ranging from 5 to 9. Figure 1 gives the distribution of optimal orders of percentiles for response cardinality 5 when \(b\) lies relatively closer to \(a\).

In Figure 1, moving from left to right, the first three boxes represent the optimal orders of percentile \(p_1\) with respect to the entropy values 0, 0.72 and 0.97. Fourth to sixth boxes represent the optimal orders of percentile \(p_2\) with respect to the entropy values 0, 0.72 and 0.97 and seventh to ninth boxes represent the optimal orders of percentile \(p_3\) with respect to the entropy values 0, 0.72 and 0.97. The boxplot is drawn based on all datasets generated from 9 intervals where the parameter \(b\) lies closer to \(a\) with response cardinality 5. Figure 1 indicates that there is no visible difference
between the optimal orders of percentiles $p_1$ as well as $p_3$ with respect to the entropy values 0.72 and 0.97. However, the difference is visible between the optimal order of percentile $p_2$ with respect to the entropy value 0 when compared to the optimal orders of $p_2$ with respect to the entropy values 0.72 and 0.97. Hence, it requires a statistical examination.

Table 2. Sample output when $b$ is closer to $a$

<table>
<thead>
<tr>
<th>Sl No.</th>
<th>Entropy</th>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$p_3$</th>
<th>$\hat{a}$</th>
<th>$\hat{b}$</th>
<th>$\hat{c}$</th>
<th>AEP</th>
<th>AEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.97</td>
<td>0.31</td>
<td>0.5</td>
<td>0.97</td>
<td>5.75</td>
<td>9.5</td>
<td>18.97</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.1</td>
<td>0.5</td>
<td>0.87</td>
<td>5.5</td>
<td>10.5</td>
<td>21.08</td>
<td>0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>3</td>
<td>0.72</td>
<td>0.31</td>
<td>0.5</td>
<td>0.85</td>
<td>5.75</td>
<td>9.5</td>
<td>21.64</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>4</td>
<td>0.97</td>
<td>0.31</td>
<td>0.5</td>
<td>0.68</td>
<td>5.75</td>
<td>9.5</td>
<td>21.33</td>
<td>0.13</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>0.97</td>
<td>0.2</td>
<td>0.51</td>
<td>0.87</td>
<td>4.55</td>
<td>9.43</td>
<td>21.55</td>
<td>0.08</td>
<td>0.18</td>
</tr>
<tr>
<td>6</td>
<td>0.97</td>
<td>0.2</td>
<td>0.7</td>
<td>0.97</td>
<td>3.69</td>
<td>10.71</td>
<td>18.93</td>
<td>0.15</td>
<td>0.25</td>
</tr>
<tr>
<td>7</td>
<td>0.72</td>
<td>0.4</td>
<td>0.69</td>
<td>0.87</td>
<td>3.5</td>
<td>11.0</td>
<td>21.00</td>
<td>0</td>
<td>0.16</td>
</tr>
<tr>
<td>8</td>
<td>0.72</td>
<td>0.2</td>
<td>0.66</td>
<td>0.87</td>
<td>5.67</td>
<td>7.72</td>
<td>21.69</td>
<td>0.13</td>
<td>0.22</td>
</tr>
<tr>
<td>9</td>
<td>0.72</td>
<td>0.31</td>
<td>0.5</td>
<td>0.85</td>
<td>5.75</td>
<td>9.5</td>
<td>21.64</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>10</td>
<td>0.97</td>
<td>0.2</td>
<td>0.57</td>
<td>0.97</td>
<td>3.26</td>
<td>11.36</td>
<td>18.89</td>
<td>0.07</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 3. Sample output when $b$ is closer to $c$

<table>
<thead>
<tr>
<th>Sl No.</th>
<th>Entropy</th>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$p_3$</th>
<th>$\hat{a}$</th>
<th>$\hat{b}$</th>
<th>$\hat{c}$</th>
<th>AEP</th>
<th>AEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.97</td>
<td>0.14</td>
<td>0.74</td>
<td>0.79</td>
<td>3.45</td>
<td>15.5</td>
<td>20.5</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.13</td>
<td>0.85</td>
<td>0.88</td>
<td>3.83</td>
<td>14.5</td>
<td>19.5</td>
<td>0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>3</td>
<td>0.97</td>
<td>0.03</td>
<td>0.83</td>
<td>0.88</td>
<td>6.03</td>
<td>15.5</td>
<td>19.25</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>0.97</td>
<td>0.13</td>
<td>0.77</td>
<td>0.89</td>
<td>3.5</td>
<td>15.5</td>
<td>19.25</td>
<td>0.13</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>0.97</td>
<td>0.14</td>
<td>0.74</td>
<td>0.79</td>
<td>3.45</td>
<td>15.5</td>
<td>20.5</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>6</td>
<td>0.72</td>
<td>0.13</td>
<td>0.85</td>
<td>0.88</td>
<td>3.83</td>
<td>14.5</td>
<td>19.5</td>
<td>0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>7</td>
<td>0.72</td>
<td>0.17</td>
<td>0.81</td>
<td>0.86</td>
<td>6.25</td>
<td>15.5</td>
<td>19.25</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>8</td>
<td>0.72</td>
<td>0.03</td>
<td>0.49</td>
<td>0.89</td>
<td>6.13</td>
<td>13.77</td>
<td>19.69</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>9</td>
<td>0.72</td>
<td>0.03</td>
<td>0.74</td>
<td>0.79</td>
<td>6.03</td>
<td>15.5</td>
<td>20.5</td>
<td>0.09</td>
<td>0.19</td>
</tr>
<tr>
<td>10</td>
<td>0.97</td>
<td>0.03</td>
<td>0.69</td>
<td>0.79</td>
<td>6.11</td>
<td>14.0</td>
<td>21.5</td>
<td>0.14</td>
<td>0.25</td>
</tr>
</tbody>
</table>
To validate the observations made from Figure 1, one way ANOVA is carried out for the optimal orders of percentiles with respect to the entropy values in all 9 intervals where the parameter $b$ lies closer to $a$ for response cardinality 5. Based on the ANOVA results, it has been found that:

- There is no significant difference between the optimal orders of percentiles $p_3$;
- There are statistically significant differences between the optimal orders of percentiles $p_1$ as well as $p_2$ with respect to the entropy values 0, 0.72 and 0.97.

Hence, Tukey’s Honest Significant Difference (HSD) test has been carried out to reach a conclusion. From the Tukey’s HSD test results, it is found that:

- There is no significant difference in the optimal orders of percentile $p_1$ used by the pairs of entropy values $\{0, 0.72\}$ and $\{0.72, 0.97\}$;
- There is significant difference in the optimal orders of the percentile $p_1$ used by the pair of entropy values $\{0, 0.97\}$;
- There is no significant difference in the optimal orders of percentile $p_2$ used by the pair of entropy values $\{0.72, 0.97\}$;
- There is significant difference in the optimal orders of the percentile $p_2$ for the pairs of entropy values $\{0, 0.72\}$ and $\{0, 0.97\}$.

Boxplot of the optimal orders of percentiles with the corresponding entropy values for response cardinality 5 from all intervals where $b$ lies closer to $c$ is given in Figure 2.

From Figure 2, it has been observed that for the cases of intervals where the parameter $b$ lies closer to $c$ having response cardinality 5, ANOVA results show no significant differences between the optimal orders of percentiles $p_1$, $p_2$, and $p_3$. 
Similar conclusions can be drawn with other cases having response cardinality ranging from 6 to 9. To get an insight on the optimal orders of percentiles, boxplots corresponding to response cardinality 8 are presented in Figures 3 and 4.

Boxplot of the optimal orders of percentiles with the corresponding entropy values for response cardinality 8 from all intervals where \( b \) lies closer to \( a \) is given in Figure 3.

Boxplot of the optimal orders of percentiles with the corresponding entropy values for response cardinality 8 from all intervals where \( b \) lies closer to \( c \) is given in Figure 4.

It may be noted that the conclusions drawn from the above analysis provide results of varying nature. In some cases, a clearcut choice of optimal orders of percentiles could be found whereas in remaining cases the findings are inconclusive. Hence, it is necessary to devise a method for finding the best set of optimal orders of percentiles \( p_1, p_2 \) and \( p_3 \) for estimating the parameters \( a, b \) and \( c \). Towards this, the optimal orders of percentiles from each of the intervals are
Figure 4. Boxplot of optimum percentile orders when \(b\) lies closer to \(c\) for response cardinality 8

![Box Plot for Optimal Percentile Order Values](image)

Grouped according to the entropy values. The frequencies of occurrence of entropy values vary among different intervals. Hence, it is suggested to identify the best optimal orders of percentiles for estimation of parameters based on the weighted averages of the optimal orders of \(p_1, p_2\) and \(p_3\). Based on the numerical study, the optimal orders of percentiles when \(b\) lies closer to \(a\) are listed in Table 4.

Optimal orders of percentiles in intervals where \(b\) lies closer to \(c\) based on the numerical study are given in Table 5.

Recommendations for best optimal orders of percentiles \(p_1, p_2\) and \(p_3\) are given based on Tables 4 and 5. In the case of intervals where the parameter \(b\) lies closer to \(a\), the best optimal orders of \(p_1\) is 0.15, \(p_2\) is 0.6 and that of \(p_3\) is 0.8 for response cardinalities ranging from 5 to 9. In the case of intervals where the parameter \(b\) lies closer to \(c\), the best optimal orders of percentiles \(p_1, p_2\) and \(p_3\) are 0.15, 0.6 and 0.8 for the response cardinalities ranging from 5 to 9.

To illustrate the methodology considered in this paper, an example has been constructed based on the data used by Wang, Gao and Guo (2012).

### 5.3. Example

Experimental data set for the average scores and belief degree levels of experts on the degree of difficulty of a question paper analysed by experts based on their experience is given below:

\[\{(39,0.32),(45,0.55),(59,0.68),(70,0.81),(95,0.98)\}\]

It is verified from the plot that the given experimental data resembles zigzag uncertainty distribution. The optimal orders of percentiles for estimating the parameters have been obtained as \(p_1 = 0.32, p_2 = 0.55\) and \(p_3 = 0.81\). The estimates obtained by using the optimal orders of percentiles are \(\hat{a} = 36.88, \hat{b} = 40.19\) and \(\hat{c} = 88.26\). Further, the estimated uncertainty distribution values have also been found as 0.32, 0.55, 0.69, 0.81 and 1. Figure 5 shows the experimental data and the fitted distribution obtained using the percentile matching method.
In this paper, estimation of unknown parameters involved in a zigzag uncertainty distribution has been carried out using the method of percentile matching and method of moments. An experimental study has been conducted using data sets simulated from 18 intervals. Intervals were chosen based on the position of the unknown parameter $b$ with respect to the parameters $a$ and $c$. Simulated datasets from all intervals have response cardinality ranging from 5 to 9.

**Table 4. Optimal orders of percentiles for intervals where $b$ lies closer to $a$**

<table>
<thead>
<tr>
<th>Response Cardinality</th>
<th>Entropy</th>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$p_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>0.12</td>
<td>0.76</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.13</td>
<td>0.58</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.13</td>
<td>0.65</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.15</td>
<td>0.55</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.25</td>
<td>0.67</td>
<td>0.86</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper, estimation of unknown parameters involved in a zigzag uncertainty distribution has been carried out using the method of percentile matching and method of moments. An experimental study has been conducted using data sets simulated from 18 intervals. Intervals were chosen based on the position of the unknown parameter $b$ with respect to the parameters $a$ and $c$. Simulated datasets from all intervals have response cardinality ranging from 5 to 9.
From the experimental study, it is found that the method of percentile matching gives better estimates than the method of moments for the parameters involved in zigzag uncertainty distribution on using the percentiles of optimal orders based on minimum absolute error. This holds true in all intervals irrespective of the position of the parameter $b$ with respect to $a$ and $c$.

The boxplot diagrams of the optimal orders of percentiles with respect to entropy values showed visible differences in some of the cases. Hence, one way ANOVA has been carried out.
for analysing the significant differences between the optimal orders of percentiles with respect to the entropy values in all intervals. Based on the ANOVA results, Tukey’s HSD test has been carried out to find the pairs of entropy values having no significant differences in the optimal orders of percentiles. The choice of optimal orders of percentiles remained inconclusive in some of the cases. Hence, weighted average of optimal orders of percentiles has been found using the frequency of occurrence of the entropy values as weights for simulated datasets in all 18 intervals.

Based on the numerical study, the recommended values for the best optimal orders of percentiles are given in Table 6.

The method of percentile matching can be applied to other uncertainty distributions as well. The experimental study clearly shows the superiority of the method of percentile matching over the method of moments. Here, the optimal orders of percentiles have been suggested based on a search procedure. It is possible to think of applying soft computing algorithms like genetic algorithm for the estimation of unknown parameters in zigzag uncertainty distribution as well as that involved in other uncertainty distributions.
REFERENCES


S. Sampath is currently holding the position of Professor of Statistics in the University of Madras, Chennai, India. He got his Ph.D. degree from the University of Madras. He has served in Loyola College, Chennai and Birla Institute of Science and Technology, Dubai, UAE before joining University of Madras in 2008. He has more than 35 years' experience in teaching and research. He has guided many scholars for their research degrees and currently four are working for their Ph.D. degree under his guidance. His main areas of research are Inference for Finite Populations, Credibility Theory, Testing and Estimation of Parameters in Imprecise Models and Data Mining. He has to his credit more than 50 research articles in highly rated research journals. His book titled “Sampling Theory and Methods” published by CRC Press and Narosa Publishing house is a popular book in the academic circles.

K. Anjana is a full-time research scholar working for her Ph.D. degree in the Department of Statistics, University of Madras, Chennai. Her main area of research is estimation of parameters in uncertainty distributions.