Chapter 2
Percentile Matching Estimation of Linear Uncertainty Distribution
Chapter 2

Percentile Matching Estimation of Linear Uncertainty Distribution

2.1 Introduction

Basic concepts in uncertainty theory have been discussed in Chapter 1 of this thesis. Descriptions about various uncertainty distributions also have been explained in the first chapter. An uncertainty distribution suitable for a given situation is mainly identified with the help of distribution of belief levels. Understanding of the shapes of distributions is insufficient to study about their properties and applications in real life situations. Conclusions based on an uncertainty distribution would be practically useful only when a complete description of the distribution is available. Only when the values of parameters (constants appearing in an uncertainty distribution) are fully known, the distribution is complete and can be used in practice under uncertain situations. Hence, in this chapter we consider the problem of estimating parameters present in linear uncertainty distribution based on experimental datasets. Experimental datasets are made up of possible values of uncertain variable along with expert’s belief levels. Following Liu (2010c), it is decided to adopt tools available in Statistical Theory of Estimation.

Several methods are available for estimating the unknown parameters of probability distributions. Method of least squares, Method of moments, 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

The contents of this chapter is available in Sampath and Anjana (2016)
probability theory, estimation techniques have also been developed in uncertainty
type. Liu (2007) introduced the concept of moments in uncertainty theory. The
study of uncertain statistics was started by Liu (2010c). Uncertain statistics refers to a
methodology used for collecting and interpreting expert’s experimental data by
uncertainty theory. Liu (2010c) applied the method of least squares in studies related
to uncertainty theory. Delphi method for estimating uncertainty distributions was
proposed by Wang, Gao and Guo (2012a). Wang and Peng (2014) proposed the
method of moments as a technique for estimating the unknown parameters of
uncertainty distributions. Liu (2015) 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, Panjer and Wilmot (2008). A brief
account on percentile matching method in the context of Statistical Theory of
Estimation is given in Section 2.2. 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 chapter, 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 percentile matching method with the existing competitors by
way of numerical studies.

This chapter is organized as follows. The second section deals with the
commonly used estimation procedures available in Statistical Estimation Theory
namely, method of moments which uses theoretical and empirical moments, method
of least squares based on the residuals and method of percentile matching which
makes use of percentiles obtained from sample observations and those based on the
assumed probability model. Third section is devoted for discussion on methods meant
for estimating unknown parameters of uncertainty distributions. Fourth section
discusses the outcome of experimental studies carried out for estimating the unknown
parameters of linear uncertainty distribution. Findings and conclusions are given in the last section.

2.2 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, method of least squares etc. 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 available in Statistical Theory of Estimation are presented in this section. Method of least squares is used for estimating parameters by minimizing the sum of squared distances 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 by Klugman, Panjer and Wilmot (2008) uses percentiles of different orders of the 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 \( nk^{th} \) observation. A percentile matching estimate of the vector valued parameter \( \theta = (\theta_1, \theta_2, ..., \theta_p) \) is a 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 \( x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)} \) are the order statistics from the sample.

It is to be noted that, \( F(\pi_{g_k} | \theta) = g_k, k = 1, 2, 3, \ldots, p \) where \( F(\cdot) \) 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 \( g < \frac{l}{(n+1)} \) 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.

Thus, brief explanations regarding the estimation methods available in Statistical Estimation Theory have been given in this section. The following section explores the application of the method of least squares, method of moments and method of percentile matching for estimating the unknown parameters of linear uncertainty distribution.

2.3 Methods for Estimating the Unknown 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**

Liu (2010c) adopted the method of least squares in uncertainty theory. Details related to the application of least squares method are furnished below. Suppose that an uncertainty distribution to be determined has a known functional form \( \Phi(x|\theta_1, \theta_2, \ldots, \theta_p) \) having parameters \( \theta_1, \theta_2, \ldots, \theta_p \). To estimate the parameters \( \theta_1, \theta_2, \ldots, \theta_p \), the method of least squares minimizes the sum of the squares of the
distances of expert’s experimental data from the uncertainty distribution. That is, for a given set of expert’s experimental data \((x_1, \alpha_1), (x_2, \alpha_2), \ldots, (x_n, \alpha_n)\), the least squares estimates of \(\theta_1, \theta_2, \ldots, \theta_p\) are found by minimizing

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

with respect to \(\theta = \theta_1, \theta_2, \ldots, \theta_p\).

While estimating uncertain parameters, it may be noted that closed form solutions for least squares 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 http://orcs.edu.cn/online/ makes the computation of least squares estimates an easy one.

**Method of moments**

Wang and Peng (2014) proposed the method of moments for estimating unknown parameters of an uncertainty distribution. The method is as follows.

Let \(\xi\) be a non-negative uncertain variable having an uncertainty distribution \(\Phi(x|\theta_1, \theta_2, \ldots, \theta_p)\) with unknown parameters \((\theta_1, \theta_2, \ldots, \theta_p)\). A set of expert’s experimental data \((x_1, \alpha_1), (x_2, \alpha_2), \ldots, (x_n, \alpha_n)\) where

\[
0 \leq x_1 < x_2 < \ldots < x_n, 0 \leq \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n \leq 1
\]

is given.

The expression for \(k^{th}\) empirical moment of \(\xi\) due to Wang and Peng (2014) obtained with the help of empirical uncertainty distribution is given by (1.10) namely,

\[
\bar{\xi}^k = \alpha_i x_i^k + \frac{1}{k+1} \sum_{i=1}^{n} \sum_{j=0}^{k} (\alpha_{i+1} - \alpha_i) x_i^j x_{i+1}^{k-j} + (I - \alpha_n) x_n^k.
\]

The moment estimates \((\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p)\) 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 given below.
\[
\int_0^{\infty} \left( 1 - \Phi \left( \frac{x}{\sqrt{k}} \theta_1, \theta_2, \ldots, \theta_p \right) \right) dx = \overline{\xi}^k, \quad k = 1, 2, 3, \ldots, p
\]

where \( \overline{\xi}^1, \overline{\xi}^2, \ldots, \overline{\xi}^p \) are the empirical moments found using (1.10).

For example, let \( \xi \) be a linear uncertain variable \( \xi \sim \mathcal{L}(a, b) \) with two unknown parameters \( a > 0 \) and \( b > 0 \) 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 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 (2015) are given by

\[
E[\xi] = \frac{a+b}{2}
\]

and

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

First and second empirical moments denoted by \( \overline{\xi}^1 \) and \( \overline{\xi}^2 \) are calculated using (1.10) by substituting \( k = 1 \) 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.

The utility of percentile matching method for estimation of parameters of uncertainty distributions is examined below. The proposed method uses the percentiles based on the empirical uncertainty distribution and the true uncertainty distribution.

**Method of percentile matching**

Consider a set of expert’s experimental data

\[(x_j, \alpha_j), (x_2, \alpha_2), \ldots, (x_n, \alpha_n)\]
where $x_i, i = 1, 2, 3, \ldots, n$ are the observed values and $\alpha_i, i = 1, 2, 3, \ldots, n$ are the belief degrees. Here, it is assumed that $0 \leq x_1 < x_2 < \ldots < x_n, 0 \leq \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n \leq 1$.

The observed values $x_i, i = 1, 2, 3, \ldots, n$ are expected to lie in the interval $(a, b)$. Using (1.3) given in the first chapter, empirical uncertainty distribution can be constructed for the expert’s experimental data. 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.

Steps involved in the percentile matching method employed to estimate the parameters in linear uncertainty distribution is given below.

Let $\xi$ be a linear uncertain variable $\xi \sim \xi \sim \mathcal{L}(a, b)$ with two unknown parameters $a > 0$ and $b > 0$ 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 $x_1 = \Phi_\alpha^{-1}(p_1)$ and $x_2 = \Phi_\alpha^{-1}(p_2)$. 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 this section, the methods for estimating the unknown parameters in uncertainty distributions namely, method of least squares, method of moments and method of percentile matching have been discussed. A detailed numerical study has been carried out on the estimation of parameters in linear
uncertainty distribution in the following section. It also gives a comparison on the performances of method of least squares, method of moments and method of percentile matching in estimating linear uncertain parameters.

2.4 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 \[ \sum_{j=1}^{n} |\hat{\Phi}(x_j) - \alpha_j| \] where \( \hat{\Phi}(x_j) \) are the estimated belief values and \( \alpha_j \) are the corresponding experimental belief values. In further discussion, it will be denoted by \( AE \). That is,

\[ AE = \sum_{j=1}^{n} |\hat{\Phi}(x_j) - \alpha_j| \] (2.1)

In the comparative study, experimental data sets which are 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 \( \Phi \) for \( x_j \)'s obtained in step (i) namely, \( \Phi(x_1), \Phi(x_2), \ldots, \Phi(x_n) \).
(iii) $\Phi(x_i)$ values obtained in step (ii) are either added with or subtracted in a randomized manner by a quantity denoted by $\varepsilon$ which is defined as follows.

$$\varepsilon = \frac{\Phi(x_2) - \Phi(x_i)}{c}$$

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,

$$\Phi(x_1) \pm \varepsilon, \Phi(x_2) \pm \varepsilon, ..., \Phi(x_n) \pm \varepsilon.$$

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

In the experimental study, for $a$ and $b$, twenty pairs of values with varying differences namely, 10, 20, 30 and 50 (each comprising five sets) as given in Table 2.1 below are used.

<table>
<thead>
<tr>
<th>Differences</th>
<th>((a,b))</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>(5,15)</td>
</tr>
<tr>
<td></td>
<td>(10,20)</td>
</tr>
<tr>
<td></td>
<td>(15,25)</td>
</tr>
<tr>
<td></td>
<td>(20,30)</td>
</tr>
<tr>
<td></td>
<td>(25,35)</td>
</tr>
<tr>
<td>20</td>
<td>(5,25)</td>
</tr>
<tr>
<td></td>
<td>(10,30)</td>
</tr>
<tr>
<td></td>
<td>(15,35)</td>
</tr>
<tr>
<td></td>
<td>(20,40)</td>
</tr>
<tr>
<td></td>
<td>(25,45)</td>
</tr>
<tr>
<td>30</td>
<td>(10,40)</td>
</tr>
<tr>
<td></td>
<td>(15,45)</td>
</tr>
<tr>
<td></td>
<td>(20,50)</td>
</tr>
<tr>
<td></td>
<td>(25,55)</td>
</tr>
<tr>
<td></td>
<td>(30,60)</td>
</tr>
<tr>
<td>50</td>
<td>(10,60)</td>
</tr>
<tr>
<td></td>
<td>(15,65)</td>
</tr>
<tr>
<td></td>
<td>(20,70)</td>
</tr>
<tr>
<td></td>
<td>(25,75)</td>
</tr>
<tr>
<td></td>
<td>(30,80)</td>
</tr>
</tbody>
</table>

**Table 2.1:** Choices of intervals

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 earlier 100 times by using randomly generated $\varepsilon$’s (described
in (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 $\min(\alpha)+0.01$ 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 $\max(\alpha)−0.0001$ are assigned to $p_2$ in step 0.01.

For each pair of $p_1$ and $p_2$, 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 with $a=5$ and $b=15$. That is, $x_i,i=1,2,...,10$ takes values in the interval $(5,15)$. The $x$ values defined by

$$a+(b-a)*0.10*x$$

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 $\epsilon$ 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 $\epsilon$ 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 estimate 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 squares 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 $\epsilon$) 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 which can be quantified using entropy of the distribution of $+\epsilon$ and $-\epsilon$. The entropy based on the distribution of $+\epsilon$ and $-\epsilon$ 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 $+\epsilon$ and $-\epsilon$ generated while simulating belief degree values. Entropy value 0 indicates that the belief values are obtained by a complete shift of $+\epsilon$ or $-\epsilon$ 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 2.2 presented in Page 31 of this thesis. Entries in a row are values of the absolute errors due to three different methods of estimation considered for the study 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 $+\epsilon$ and $-\epsilon$ in the simulation process. From Table 2.2, 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 squares 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 are 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 \(p_2\) 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 Figures 2.1 and 2.2 respectively presented in Pages 32 and 33 of this thesis. 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 has been found that differences between the \(p_1\) values are statistically significant with respect to the
<table>
<thead>
<tr>
<th>Sl No.</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>0.9710</td>
<td>0.0704</td>
<td>0.1304</td>
<td>5.2963</td>
<td>15.2963</td>
<td>0.2370</td>
<td>0.2430</td>
<td>0.2615</td>
</tr>
<tr>
<td>2</td>
<td>0.9710</td>
<td>0.3104</td>
<td>0.6604</td>
<td>4.3792</td>
<td>15.4247</td>
<td>0.2124</td>
<td>0.3025</td>
<td>0.2485</td>
</tr>
<tr>
<td>3</td>
<td>0.9710</td>
<td>0.0704</td>
<td>0.8404</td>
<td>5.3505</td>
<td>14.5809</td>
<td>0.1259</td>
<td>0.1462</td>
<td>0.1271</td>
</tr>
<tr>
<td>4</td>
<td>0.9710</td>
<td>0.1596</td>
<td>0.8396</td>
<td>5.4344</td>
<td>14.5694</td>
<td>0.2287</td>
<td>0.2322</td>
<td>0.2485</td>
</tr>
<tr>
<td>5</td>
<td>0.9710</td>
<td>0.2504</td>
<td>0.7204</td>
<td>5.2963</td>
<td>15.2963</td>
<td>0.2370</td>
<td>0.2660</td>
<td>0.2787</td>
</tr>
<tr>
<td>6</td>
<td>0.9710</td>
<td>0.1296</td>
<td>0.4196</td>
<td>4.5865</td>
<td>15.4907</td>
<td>0.2149</td>
<td>0.3213</td>
<td>0.2327</td>
</tr>
<tr>
<td>7</td>
<td>1.0000</td>
<td>0.4996</td>
<td>0.7796</td>
<td>4.4754</td>
<td>15.5283</td>
<td>0.1358</td>
<td>0.2983</td>
<td>0.1365</td>
</tr>
<tr>
<td>8</td>
<td>0.8813</td>
<td>0.3904</td>
<td>0.9204</td>
<td>4.7037</td>
<td>14.7037</td>
<td>0.1778</td>
<td>0.2222</td>
<td>0.2198</td>
</tr>
<tr>
<td>9</td>
<td>0.8813</td>
<td>0.0704</td>
<td>0.3604</td>
<td>5.2963</td>
<td>15.2963</td>
<td>0.1778</td>
<td>0.2680</td>
<td>0.2485</td>
</tr>
<tr>
<td>10</td>
<td>0.8813</td>
<td>0.1296</td>
<td>0.6996</td>
<td>4.5689</td>
<td>15.6086</td>
<td>0.1527</td>
<td>0.2758</td>
<td>0.1573</td>
</tr>
</tbody>
</table>

*Table 2.2: Sample output*
Figure 2.1: Boxplot of distribution of percentiles for data sets with 5 values.
Figure 2.2: Boxplot of distribution of percentiles for data sets with 10 values
<table>
<thead>
<tr>
<th>Number of values</th>
<th>Entropy</th>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$p_1$</th>
<th>$p_2$</th>
<th>$p_1$</th>
<th>$p_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.7219</td>
<td>0.1617</td>
<td>0.3461</td>
<td>0.9710</td>
<td>0.1586</td>
<td>0.6188</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.6500</td>
<td>0.1670</td>
<td>0.3075</td>
<td>0.9183</td>
<td>0.1839</td>
<td>0.5170</td>
<td>0.0000</td>
<td>0.1586</td>
<td>0.7195</td>
</tr>
<tr>
<td>7</td>
<td>0.5917</td>
<td>0.1622</td>
<td>0.3315</td>
<td>0.8631</td>
<td>0.1820</td>
<td>0.4483</td>
<td>0.9852</td>
<td>0.1561</td>
<td>0.6552</td>
</tr>
<tr>
<td>8</td>
<td>0.5436</td>
<td>0.1540</td>
<td>0.3131</td>
<td>0.8113</td>
<td>0.1783</td>
<td>0.3947</td>
<td>0.9544</td>
<td>0.1561</td>
<td>0.6965</td>
</tr>
<tr>
<td>9</td>
<td>0.5033</td>
<td>0.1329</td>
<td>0.2801</td>
<td>0.7642</td>
<td>0.1563</td>
<td>0.3375</td>
<td>0.9183</td>
<td>0.1755</td>
<td>0.6599</td>
</tr>
<tr>
<td>10</td>
<td>0.4690</td>
<td>0.1213</td>
<td>0.2675</td>
<td>0.7219</td>
<td>0.1557</td>
<td>0.3118</td>
<td>0.8813</td>
<td>0.1772</td>
<td>0.6963</td>
</tr>
</tbody>
</table>

Table 2.3: Recommended percentile values
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 Significant Difference (HSD) test has been carried out to reach conclusive evidence. It was observed that

\[(0.72,0.97),(0.65,1),(0.59,0.98),(0.81,0.95) \text{ and } (0.91,0.99)\]

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 2.3 presented in Page 34 of this thesis 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.

In this section, a detailed experimental study has been conducted on estimating the unknown parameters involved in linear uncertainty distribution. A comparative study on the performances of the method of least squares, method of moments and method of percentile matching has been carried out in this section. The findings and conclusions are presented in the following section.

2.5 Summary

In this chapter, 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 2.3 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.