Building Statistical Hypothesis Tests for Fuzzy Data and Their Applications to Decision Making

Pei-Chun LIN
Graduate School of Information, Production and Systems
Waseda University

April 2013
## Contents

Title Page          i

Contents          ii

Abstracts          v

Citations to Previous Publish Work          x

List of Figures          xi

List of Tables          xii

Acknowledgements          xiv

Dedications          xvi

1 Introduction          1
  1.1 Background          1
  1.2 Motivation and Objective          3
  1.3 Research Framework          4
  1.4 Structure of the Thesis          4

2 Some Basic Theoretical Concepts          8
  2.1 Fuzzy Numbers          8
  2.2 Fuzzy Statistic Analysis          10
  2.3 Nonparametric Statistical Method          10
    2.3.1 Kolmogorov-Smirnov Two-Sample Test          10
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Kolmogorov-Smirnov Two Sample Test with Continuous Fuzzy Data</td>
<td>13</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>13</td>
</tr>
<tr>
<td>3.2</td>
<td>Preliminary Preparation</td>
<td>14</td>
</tr>
<tr>
<td>3.3</td>
<td>Kolmogorov-Smirnov Two-Sample Test with Continuous Fuzzy Data</td>
<td>15</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Empirical distribution function with continuous fuzzy data</td>
<td>15</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Kolmogorov-Smirnov two-sample test with continuous fuzzy data</td>
<td>19</td>
</tr>
<tr>
<td>3.4</td>
<td>Empirical Studies</td>
<td>20</td>
</tr>
<tr>
<td>3.5</td>
<td>Summary of Chapter 3</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>Identifying the Distribution Differences of Fuzzy Data Based on a</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Nonparametric Statistical Method</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction and Literature Review</td>
<td>28</td>
</tr>
<tr>
<td>4.2</td>
<td>Identifying the Distribution Difference of Fuzzy Data Based on a</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Nonparametric Statistical Method</td>
<td></td>
</tr>
<tr>
<td>4.2.1</td>
<td>Realization of a Continuous Fuzzy Data</td>
<td>30</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Identifying the Distribution Difference of Fuzzy Data Based on a</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Nonparametric Statistical Method</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Empirical Studies</td>
<td>37</td>
</tr>
<tr>
<td>4.4</td>
<td>Comparison of Each Methods</td>
<td>42</td>
</tr>
<tr>
<td>4.5</td>
<td>Summary of Chapter 4</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>Risk Assessment of a Portfolio Selection Model Based on a Fuzzy</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Statistical Test</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>47</td>
</tr>
<tr>
<td>5.2</td>
<td>Notations and Preliminary Definitions</td>
<td>49</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Statistical Analysis</td>
<td>51</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Markowitz’s Portfolio Selection Model</td>
<td>52</td>
</tr>
<tr>
<td>5.3</td>
<td>Fuzzy Statistical Test on the Portfolio Selection Model</td>
<td>53</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Portfolio Selection Model with Interval Values</td>
<td>54</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Fuzzy Statistical Test for the Portfolio Selection Model</td>
<td>56</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Procedure of Solving Portfolio Selection Model with Interval Values</td>
<td>57</td>
</tr>
<tr>
<td>5.4</td>
<td>Empirical Studies</td>
<td>58</td>
</tr>
<tr>
<td>5.5</td>
<td>Summary of Chapter 5</td>
<td>64</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Discussion</td>
<td>64</td>
</tr>
</tbody>
</table>

iii
## 6 A Parametric Assessment Approach to Solving Facility Location Problems with Fuzzy Demands

6.1 Introduction and Literature Review .............................................. 68  
6.2 Preliminary Definitions ................................................................. 71  
6.3 Problem Statements ................................................................. 74  
6.4 Empirical Study ................................................................. 80  
6.5 Summary of Chapter 6 ............................................................... 86  

### 6.5.1 Comparison ................................................................. 86  
### 6.5.2 Conclusions ............................................................... 88  

## 7 Conclusions and Future Work

7.1 Conclusions ................................................................. 90  
7.2 Future Work ................................................................. 92  

### Bibliography

94  

### Appendix of Tables

100  

Appendix A Kolmogorov-Smirnov Two-Sample Statistic .................. 100  

### List of Publications

101
Abstract

In conventional statistical methods, hypothesis tests play a fundamental role in making decisions. But in real-world applications, sometimes vague information is given such as in linguistic expression like “parts of the product are good.” It is not easy to deal with such linguistic expressions in statistical terms. Therefore, we must establish some statistical methods to deal with those vague data.

A statistical hypothesis test plays a pivotal role in social science research. This test analyzes data in either a controlled experiment or an observational study (not controlled) for making decisions. Relevant null hypothesis $H_0$ and alternative hypothesis $H_1$ are stated the first step in testing process. We should decide which test should be used and select a significance level $\alpha$ for making decision. Popular significance levels are 10%, 5%, 1%, 0.5%, and 0.1%. Usually, when $\alpha$ is chosen, we have the $(1 - \alpha)$ confidence interval for the parameter which we want to estimate. We do not reject the $H_0$ when the statistic value is included in $(1 - \alpha)$ confidence interval. The statistic test results under a pre-specified significance level can help us to decide whether experimental results contain enough information to cast doubt on conventional perception.

The objective of this thesis is to create a new type of fuzzy hypothesis test that can deal with continuous fuzzy data. In addition, we also explain various statistically significant results in risk and error assessment applications.

In the thesis, extending conventional statistical method, we build up new statistical methods that can deal with fuzzy numbers. We call this type of statistical method as ”fuzzy statistics.” Following Zadeh’s concepts and definitions, we use fuzzy set theory to deal with the fuzzy statistics.
in the thesis. This thesis focuses on two topics. First, we build a new nonparametric statistical test for fuzzy data that can identify distribution differences. Second, a new statistical hypothesis test will be proposed for fuzzy data. In addition, this thesis illustrates two real-life applications of fuzzy statistical test. The structure of the thesis is summarized as follows:

Chapter 1 of this thesis provides the background and motivation for the study as well as the objective of this thesis. We also describe the framework and structure of this thesis.

Chapter 2 provides some preliminary concepts and methods, including fuzzy numbers, fuzzy statistical analysis and a nonparametric statistical method.

In chapter 3, we extend the Kolmogorov-Smirnov (K-S) two-sample test for continuous fuzzy data. The K-S two-sample test is a goodness-of-fit test that is used to determine whether two underlying one-dimensional probability distributions differ. To find the statistical value of a K-S two-sample test, we calculate the cumulative distribution function by means of the empirical distribution function. In this chapter, we define a new function called the weight function (denoted as WF) that can defuzzify the continuous fuzzy data into real numbers. Thus, the empirical distribution function can be estimated by using those real numbers obtained from defuzzification. In this chapter, we treat three types of fuzzy data in empirical studies. That is, we handle interval values, triangular fuzzy data or trapezoidal fuzzy data in K-S two-sample test. We also provide various significant levels $\alpha$ in this chapter to indicate different results in using K-S two-sample test for continuous fuzzy data. When we used triangular fuzzy data or trapezoidal fuzzy data for K-S two-sample test, we obtained the same statistic value 0.3 in 80%, 90%, 95%, 98% and 99% confidence interval in our method and conventional method, where the conventional method used central points to defuzzify the fuzzy data and used this defuzzification in K-S two-sample. Moreover, when we used interval value for K-S two-sample test, we obtained 0.5 for the statistic value all in 90%, 95%, 98% and 99% confidence interval in our method and we obtained 0.8 for the statistic value that is only in
99% confidence interval in conventional method. It means that we need stronger evidence to confirm the hypothesis when we used interval values for K-S two-sample test in conventional method. Hence, we conclude that our method is more extensive to use K-S two-sample test for continuous fuzzy data that can enable us to judge whether or not two independent samples of continuous fuzzy data come from the same population.

Continuously, we discuss the K-S two-sample test in chapter 4. In this chapter, we compare our proposed method with various methods in identifying the probability distribution differences between two populations of fuzzy data. We derive a function, called realization of a continuous fuzzy data (RF) that can defuzzify continuous fuzzy data. The function RF is different from the function WF in chapter 3. The function WF considers a random variable $k$, central point and radius but in chapter 4 we consider only the central point and radius. The K-S two-sample test is also used in this chapter for distinguishing two populations of fuzzy data. We illustrated four different defuzzification methods for K-S two-sample test in empirical studies. We proposed a ranking criterion of function RF in this chapter. We said that the fuzzy data are in the same class if they have the same value for defuzzification; otherwise, they are in difference classes. We use function RF to defuzzify the fuzzy data and calculate the empirical distribution function by those defuzzifications. We obtained 0.3 for the statistic value that is in 95% confidence interval in four different defuzzification methods in the experiment and obtained 19 classes in our method (RF method). This number 19 is more than the number of classes, which is 18, in conventional method. Moreover, we have proved that the function WF is a decreasing function. Function WF for K-S two-sample test in chapter 3 does not satisfy the ranking criterion which we proposed in this chapter. Hence, it can be concluded that the proposed function RF in chapter 4 is successful in distinguishing two populations of continuous fuzzy data.

In chapter 5, we apply a $t$-test of fuzzy data to evaluate different risks in a portfolio selection model with fuzzy data. The central points and radiuses of fuzzy numbers are used to solve the portfolio selection problem.
We statistically evaluate the expected return with different risks by using $t$-test. Empirical studies are presented to illustrate the risk evaluation of the portfolio selection model for interval values, which was proposed by using central point and radius. We provided different risks $k$, which is a restriction for variance calculated by radius, for investors to make decision. The results of portfolio selection model were interval values composed of central points and radiuses. We obtained the results that we had a stable expected return $[5.50, 6.56]$ because we had the same expected return of radius when $k \geq 2.81$ in our proposed model. Moreover, we obtained a negative value of expected return when $k \leq 2$ in our proposed model. An investor could consider to buy a portfolio when the value $k > 2$ because we did not want to buy a portfolio with negative expected return in this experiment. Comparing with other researcher’s method (Zhang implemented the concept of the $\gamma$-level to deal with the optimization model), we obtained the expected return $[5.41, 6.43]$ under 95% confidence interval by using the same data in our experiment. The expected return of Zhang’s method is less than our expected return. We concluded that the fuzzy statistical test enables us to evaluate a stable expected return and low-risk investment under different choices of $k$.

In chapters 3 to 5, we proposed the statistical methods for fuzzy data. In chapter 6, we describe a real-world application that combines the concept of fuzzy statistics with error assessment. In real-world applications, sometimes randomness and fuzziness may coexist. In facility-location problems, vague information included in linguistic should be analyzed. We discussed uncertain demands, called fuzzy demands, in facility location problem. In the facility location model, the parameters of a fuzzy demand are determined by calculating the estimated expected value (EE value) of the fuzzy demand. It was obtained by using estimated parameters of the underlying probability distribution function of the fuzzy data. We proposed a defuzzification formula for the fuzzy demand, called the realization of the fuzzy demand ($RFD$). The RFD formula comprised the upper bound of the fuzzy demand ($RFD^+$) and the lower bound of the fuzzy demand ($RFD^-$).
Moreover, concerning the fuzzy demand, an error assessment itself was evaluated as mean absolute percentage error of the fuzzy demand (MAPE-FD). Empirical studies show that we obtained a maximum profit about 5.759 million NTD. We had less percentage error (27.13%) with respect to distance method, which was proposed by Cheng. If we did not consider MAPE-FD in RFD formula, the percentage error will increase 90.70% more from 27.13% more. But the conventional method obtained a maximum profit about 2.111 million NTD and had 53.40% percentage error with respect to distance method. The results show that, it is better to solve the real-life location problem considering the error assessment (MAPE-FD) in RFD formula.

In the final chapter, we conclude the thesis and suggest several research directions for future work. In this thesis, we have established the statistical test of fuzzy data that is called a fuzzy statistical test. We introduce a concept for "defuzzifying" fuzzy data into real numbers; that is, we use the central point and radius instead of the fuzzy data. The central point and radius will play a role as statistical characteristics similar to mean and variance. Thus, conventional statistical tests can be applicable to the parameters. To illustrate the efficacy of the proposed method, we introduce two real-life applications: one is a portfolio selection problem, and the other is a facility location problem. Our empirical studies show that we can provide various risk levels and expected returns in portfolio selection problems and it can give more choices for investors to make decisions when they buy exchange currencies. We could also achieve higher profits using the RFD formula in facility location problems. We introduce fuzzy statistical tests to deal with real-life problems in this thesis. We work through problems with interval value and triangular fuzzy data in real-life applications. If we can use fuzzy statistical tests with trapezoidal fuzzy numbers or other types of fuzzy numbers in the future, then this approach will make the proposed method more realistic. Additionally, although we can evaluate our results using a fuzzy statistical test, we also need to consider financial reports, experts’ individual experiences and other factors in the real world for a more well-rounded evaluation.
Citations to Previous Publish Work

The contents of Chapter 3 have appeared in the following published papers.


The contents of Chapter 4 are based on the following published papers.


Large portions of Chapter 5 came from the following published papers.


Finally, the contents of Chapter 6 have appeared in the following conference paper.

List of Figures

1.1 Research Framework ........................................ 5

3.1 A Triangular Fuzzy Number $f(x)$ with Central Point $o$ and Radius $l$
Have The Same Area as $h(x)$ ..................................... 17

4.1 A Trapezoidal Fuzzy Number $f(x)$ with Central Point $o$ and Radius $l$
Have The Same Area as $h(x)$ ..................................... 31

5.1 Scatterplot of $m_{o0}, m_{l0}$ vs. $k$ .......................... 63

6.1 The Potential Facility Sites and Universities in Taipei City .... 81
# List of Tables

3.1 The Price which will be Acceptable by Males and Females ........................................... 20
3.2 The Weight Values and Classes ......................................................................................... 21
3.3 The Cumulative Distributions of $X_i$ and $Y_j$ ................................................................. 21
3.4 The Price which will be Acceptable by Males and Females ........................................... 22
3.5 The Weight Values and Classes ......................................................................................... 23
3.6 The Cumulative Distributions of $X_i$ and $X_j$ ................................................................. 23
3.7 The Price which will be Acceptable by Males and Females ........................................... 24
3.8 The Weight Values and Classes ......................................................................................... 25
3.9 The Cumulative Distributions of $X_i$ and $Y_j$ ................................................................. 25
3.10 The Results of Hypothesis under Different Significant Level $\alpha$ ................................. 27

4.1 The Numbers of E-mails Per Day by Males and Females ................................................. 37
4.2 Values for $a_i$, $l_i$, $RF$ and $C_i$ of $RF$ ........................................................................ 38
4.3 The Cumulative Distributions of $X_i$ and $Y_j$ ................................................................. 38
4.4 Values for $a_i$, $l_i$, $WF$ and $C_i$ of $WF$ ........................................................................ 40
4.5 The Cumulative Distributions of $X_i$ and $Y_j$ ................................................................. 40
4.6 Values for $C_i$ and $C_i$ of $C$ ........................................................................................ 41
4.7 The Cumulative Distributions of $X_i$ and $Y_j$ ................................................................. 41
4.8 Values for $\bar{x}_i$, $\bar{y}_i$, $RD$ and $C_i$ of $RD$ ................................................................. 43
4.9 The Cumulative Distributions of $X_i$ and $Y_j$ ................................................................. 43
4.10 Values for $RF$, $C_i$ of $RF$, $WF$, $C_i$ of $WF$, $C$, $C_i$ of $C$, $RD$ and $C_i$ of $RD$ ........ 45
4.11 The Results of K-S Two Sample Test in Each Method .................................................... 45

5.1 Interval Values of Each Exchange Currency ................................................................. 58
5.2 Interval Returns of Each Exchange Currency ................................................................. 59
5.3 Central Point \( o \) and Radius \( l \) of Each Interval Values \([A, B]\) \hspace{1cm} 59
5.4 Parameters of Probability Distribution Functions for Interval Values \hspace{1cm} 60
5.5 Expected Values and Variances for Interval Values \hspace{1cm} 60
5.6 Fuzzy Statistical Test for the Results of Model (5.13) with Different Conditions \hspace{1cm} 61
5.7 The Fuzzy Expected Return with Different Conditions \( k \) \hspace{1cm} 62
5.8 Fuzzy Statistical Test for The Results of Model (5.15) and (5.16) with Different Conditions \hspace{1cm} 64

6.1 Triangular Fuzzy Numbers \((a^k, b^k, c^k)\) from Five KFCs. \hspace{1cm} 81
6.2 Three Parameters \((b^k, L^k_1, L^k_2)\) of Triangular Fuzzy Numbers from 5 KFCs. 82
6.3 The Probability Distribution Functions for Parameters \(b_j, L_{1j}, \text{ and } L_{2j}\). \hspace{1cm} 82
6.4 The EE Value of \(b_j, L_{1j}, \text{ and } L_{2j}\). \hspace{1cm} 83
6.5 Decision by \(t\)-test \hspace{1cm} 83
6.6 Percentage Error of \(EE(o_j)\) and \(EE(l_j)\) \hspace{1cm} 83
6.7 The Parameter Values of RFD \hspace{1cm} 84
6.8 The Fixed Cost, Variable Cost and Capacity for Each Potential Site. \hspace{1cm} 85
6.9 The Parameter Values of Distance Method. \hspace{1cm} 86
6.10 Solution Comparisons of Location Problem \hspace{1cm} 88
Acknowledgements

Completing my Ph.D degree is probably the most challenging activity of my life. The best and worst moments of my doctoral journey have been shared with many people. It has been a great privilege to spend four years in the Graduate School of Information, Production and Systems (IPS) at Waseda University. I am indebted to many people for making the time working with me. It’s an unforgettable experience and all members and staffs in Waseda University will always remain dear to me.

Foremost, I would like to express my sincere gratitude to my advisor Professor Junzo Watada for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. He has been a steady influence throughout my Ph.D. career.

Besides my advisor, I would like to thank the rest of my thesis committee members, Professor Hee-Hyol Lee, and Professor Tomohiro Murata, for their support, guidance and helpful suggestions.

My sincere thanks also goes to Professor Berlin Wu for providing me many helpful advices and spending his precious time to discuss with me. To work with him has been a real pleasure to me, with heaps of fun and excitement. He made me feel a friend, which I appreciate from my heart.

I especially thank all members in Watada lab, IPS, Waseda University: Nureize Bonti Arbaiy, Haydee Rocio Melo Cisneros, Yu-Yi Chu, Hao-Chen Ding, Shinya Imai, Yung-Chin Hsiao, Kim Ikno, Lee-Chuan Lin, You Li, Shamshul Bahar Yaakob, Rozlini Mohamed, Azizul Azhar Bin Ramli, Wen Song, Bo Wang, Shuming Wang, Zhen-Yuan Xu, Jian-Xiong Yang, et al. They provided many stimulating and entertaining experiences throughout
my Ph.D study. I could not complete my work without invaluable friendly assistance. I owe them my heartfelt appreciation.


My hearty thanks also go to the members in Kitakyushu Bible Church: Takemoto Kou sensei, Jessica Kondou, Kondo San, Saori Kondo, Rocky Ayatsuka, Marla Rudd Ayatsuka, Riz Crescini, Anne Larson Crescini, Itou Chiaki, Emily Ark, Junko Fukuda, Nobuko Mitani, et al. They have sheltered me over the years. They have been like surrogate families, bearing the brunt of the frustrations, and sharing in the joy of the successes. Thank you for just being there for me.

I wish to express my sincere appreciation for the financial support I received that enable me study and live for years in Japan. Sources included the Scholarship for Young Doctoral Students in Waseda University, Scholarship for Kitakyushu Science and Research Park (FAIS scholarship) and Honors Scholarship for Privately Financed International Students of Japan Student Services Organization (JASSO). Moreover, I am deeply grateful to Professor Satoshi Goto, the leader of "Ambient SoC Global COE Program of Waseda University" (GCOE) and all staffs of this program. Without sufficient financial supports from GCOE, it would have been impossible for me to attend those worldwide conferences and high level lectures.

Last but not the last, I would like to thank my love family: my parents provided a loving environment for me. My sister and my younger brother are always unconditionally support me. Their love provided my inspiration and spirituality throughout my life. I owe them a lot.

Finally, I would like to thank Tek-Min Gan. I would not complete this road without your infinite support. I hope that this work makes you proud.
Dedicated to my love family:
my father Ying-Chieh Lin,
my mother Su-Chen Chen,
my sister Mei-Ling Lin,
and my younger brother Shih-Hao Lin.
1

Introduction

1.1 Background

In social science research, many decisions, evaluations and psychological tests are conducted using surveys and/or questionnaires to seek people's opinions. It is routine to ask people about their opinions according to binary, multiple-choice questions, but people actually have complex and/or vague thoughts. If we want to understand human thinking in reality, we must create a fuzzy questionnaire (a questionnaire to collect fuzzy data) to seek people's actual thoughts. However, we seldom use fuzzy surveys (surveys using fuzzy questionnaires) in research studies because it is difficult to find an appropriate statistical method to analyze those fuzzy data.

In real-world applications, sometimes vague information is given when describing data in natural language. Knowing the probability distribution function of fuzzy data plays a pivotal role in dealing with problems in the real world. Conventional research studies in the past have not recognized the underlying probability distribution function of fuzzy data in their problems. The probability distribution function must be predicted under a specified condition or for a situation given in advance (see (50)). When we want to work with fuzzy data, the underlying probability distribution of the fuzzy data is not known. It is not easy to describe such information in statistical terms. We must establish techniques to handle such information. Following Zadeh (see (83) and (84)), we will use fuzzy set theory and take the concept of fuzzy statistics into consideration.

In conventional statistical methods, nonparametric statistical tests are a distribution-free method in that they make no assumption that the data are drawn from a particular
1. INTRODUCTION

probability distribution. The two-sample test is one of the most useful nonparametric methods for comparing two samples because it is sensitive to the differences between empirical cumulative distribution functions with regard to both the location and the shape of the two samples. Other nonparametric statistical tests may also be useful (6), such as the median test, the Mann-Whitney test and the parametric $t$-test. Although these tests are sensitive to differences between two means or medians, they can not detect any other type of difference, such as a difference in variance. One of the advantages of two-tailed tests is that such tests consistently reflect all of the types of differences between two distribution functions. The Kolmogorov-Smirnov (K-S) two-sample test is a goodness-of-fit test used to determine whether the two underlying distributions of samples differ.

To manipulate continuous fuzzy data using the K-S two-sample test, we need to calculate the empirical distribution function of the continuous fuzzy data first. Therefore, a method is necessary for classifying all of the continuous fuzzy data.

Many research works have proposed various ranking methods to classify fuzzy data. For instance, Lee-Kwang and Lee (31) proposed a method that derives rankings by considering the overall possibility distributions of fuzzy numbers and provides users with a method for evaluation. Tseng and Klein (68) designed an algorithm to rank any amount of fuzzy numbers. Ota et al. (53) developed a variable axis method (VAM) to decide the complete ordering of fuzzy numbers. Xu and Sasaki (81) proposed a vertex method to calculate the distance between Grey numbers. Lee and You (35) proposed a ranking method that generates possible ranking sequences of given fuzzy numbers. Kang et al. (21) developed a new fuzzy ranking model based on user preferences. Hung et al. (19) provided a novel accuracy function to evaluate interval-valued fuzzy information based on intuition. Moreover, Yager (82) proposed a method of ranking fuzzy numbers using a centroid index.

Fundamental statistical measurements such as the mean, the median and the mode are useful for illustrating the characteristics of a sample distribution. More research should focus on the fuzzy statistical aspects of models and their applications in engineering, medical and social science. Wu and Cheng (78) identified a model structure through qualitative simulation; Casalino et al. (5), Esogbue and Song (11), and Wu and Sun (79) discussed the concepts of fuzzy statistics and applied them to social surveys. Chen and Klein (7) proposed an approach using defuzzification methods for the fuzzy
1.2 Motivation and Objective

MADM. Wu and Tseng (80) used the fuzzy regression method of coefficient estimation to analyze the Taiwan monitoring index of economics. In addition, Wu and Sun (79) presented a set of real-life situations in which fuzzy techniques can be naturally reformulated in statistical terms. These studies have addressed various problems using defuzzification techniques to choose the central points of fuzzy numbers. Recently, Wu and Chang (77) evaluated the mean and variance values of interval data based on central point and radius data. Lin et al. (40) proposed a new weight function of fuzzy numbers defined by the central point and radius. Moreover, Lin et al. (37) proposed a method for recognizing the underlying distribution function using its central point and radius, thereby providing more information about the original fuzzy data. It is more effective to analyze the original fuzzy data. We will take this concept into consideration and integrate it with the concept of fuzzy statistics in this thesis.

In this thesis, we concentrate our discussion on fuzzy statistical tests. We developed two statistical tests of fuzzy data: a K-S two-sample test of fuzzy data, and a $t$-test of fuzzy data.

Although many papers have discussed the powerful K-S two-sample test (see the discussion in (9), (10), and (60)), these reports have all simulated them under known distributions. No statistical method can distinguish two populations of continuous fuzzy data based on their respective distribution functions. Hence, we use the K-S two-sample test to decide whether the two independent samples of continuous fuzzy data are derived from the same population. We consider a sample of continuous fuzzy data to be a set of data obtained from a single population. Given two different samples of continuous fuzzy data, our goal is to test whether they have been drawn from the same population. This method is useful in various applications, such as industry, engineering, and social surveys.

To use the K-S two-sample test, we need to calculate the empirical distribution function of continuous fuzzy data. We propose a ranking method for fuzzy data in our thesis. This method can classify all continuous fuzzy data and enable us to calculate the empirical distribution function. Although various methods have been proposed to rank fuzzy numbers, all of these methods are based on the concept of a central point.
1. INTRODUCTION

All of these methods thus ignore some information about the continuous fuzzy data in the calculation. Hence, we propose using two parameters, the central point and radius, to more effectively analyze original fuzzy data.

We still need to make some technical calculations in the processes of developing fuzzy statistical methods. In the real-life applications, we need to find out the probability distribution function of the fuzzy data and each parameter in the probability distribution functions, which enable us to calculate the values in the portfolio selection model and the facility location model. Most studies have not considered any type of probability distribution function with fuzzy data. Moreover, no statistical tests have been applied to examine the results of the proposed models. In view of these weaknesses, we have developed a $t$-test to evaluate the results (fuzzy data) of the portfolio selection model. Furthermore, an error assessment, called the mean absolute percentage error of fuzzy demand (MAPE-FD), is proposed in facility location model.

1.3 Research Framework

We show our research framework in Figure 1.1.

1.4 Structure of the Thesis

The main work of this dissertation is organized into four relatively independent parts, including (i) The Kolmogorov-Smirnov two-sample test with continuous fuzzy data, (ii) identifying the distribution differences between two populations of fuzzy data based on a nonparametric statistical method, (iii) risk assessment of a portfolio selection model based on a fuzzy statistical test, and (iv) a parametric assessment approach to solving facility location problems with fuzzy demands.

The first part of this thesis provides the background of the study as well as the motivation and objective for this thesis. Moreover, we also provide the framework for and the structure of the thesis.

In Chapter 2, we give some basic theoretical concepts that we will use in the following section including fuzzy set theory, fuzzy statistical analysis, and nonparametric statistical methods.
1.4 Structure of the Thesis

Figure 1.1: Research Framework
1. INTRODUCTION

In chapter 3, a nonparametric statistical method is discussed. We introduce the K-S two-sample test with continuous fuzzy data. The K-S two-sample test is a goodness-of-fit test that is used to determine whether two underlying one-dimensional probability distributions differ. To find the statistical pivot of a K-S two-sample test, we calculate the cumulative function by means of the empirical distribution function. When we address fuzzy data, it is essential to know how to find the empirical distribution function for continuous fuzzy data. In this chapter, we define a new function, the weight function, that can be used to address continuous fuzzy data. Moreover, we can divide samples into different classes. The cumulative function can be calculated with those divided data. The paper explains that the K-S two-sample test for continuous fuzzy data can make it possible to judge whether two independent samples of continuous fuzzy data come from the same population. The results show that it is realistic and reasonable to use the K-S two-sample test with continuous fuzzy data in social science research.

In chapter 4, we continue to introduce a nonparametric statistical method for analyzing fuzzy data. Nonparametric statistical tests are distribution-free methods without any assumption that data are drawn from a particular probability distribution. In this chapter, to identify the distribution differences between two populations of fuzzy data, we derive a function that can describe continuous fuzzy data. The function is different from the weight function in chapter 3. The K-S two-sample test is also used in this chapter for distinguishing two populations of fuzzy data. Empirical studies illustrate that the K-S two-sample test enables us to judge whether two independent samples of continuous fuzzy data are derived from the same population. The results show that the proposed function is successful in distinguishing two populations of continuous fuzzy data and is useful in various applications.

Continuing the discussion, we introduce a statistical test for fuzzy data in Chapter 5. The objective of this chapter is to develop a statistical test that can evaluate the different risks of a portfolio selection model using fuzzy data. The central points and radiuses of fuzzy numbers are used to determine the portfolio selection model, and we statistically evaluate the best return by using a fuzzy statistical test. Empirical studies are presented to illustrate the risk evaluation of the portfolio selection model with interval values. We conclude that the fuzzy statistical test employed enables us to evaluate a stable expected return and low-risk investment with different choices for $k$. 
which indicates the risk level. The results of numerical examples show that our method is suitable for short-term investments.

Thus, in chapters 3 to 5, we discuss some statistical methods for fuzzy data. There are still minor problems in dealing with fuzzy data based on fuzzy statistical tests. We introduce error assessment in Chapter 6. In real-world applications, sometimes randomness and fuzziness may coexist within a dataset. In facility location problems, data expressed in natural language contain vague information. We discuss uncertainty of the demands in facility location problems. The uncertain demands are referred to as fuzzy demands in this chapter. In the facility location model, the parameters of fuzzy demands are determined by calculating the estimated expected value (EE value) of the fuzzy demand, which is obtained by using estimated parameters of the underlying probability distribution function of the fuzzy data. Moreover, we propose a defuzzification formula for the fuzzy demand called a realization of the fuzzy demand (RFD). The RFD formula consists of the upper bound of the fuzzy demand \( RFD^+ \) and the lower bound of the fuzzy demand \( RFD^- \). Moreover, the error of a fuzzy demand is assessed as its mean absolute percentage error of the fuzzy demand (MAPE-FD). Our empirical studies show that we can solve real-life location problems by using the RFD formula and can therefore achieve higher profits in our facility location model in comparison with conventional methods.

In the last section, we conclude chapters 3 to 6 and suggest some research directions for future work.
Some Basic Theoretical Concepts

The focus in this chapter is on some basic theoretical concepts which are necessary for the discussed topic in this thesis.

2.1 Fuzzy Numbers

A fuzzy number is an extension of a regular number in the sense that it does not refer to one single value but rather to a connected set of possible values, where each possible value has its own weight between 0 and 1. This weight is called the membership function. Here, we introduce some definitions with respect to membership functions and fuzzy numbers which we will use in the following chapter.

Definition 2.1 Trapezoidal Membership Function (26)

One class of function frequently used to represent linguistic terms is the class of trapezoidal membership functions \( \mu(x; a, b, c, d) \), which are defined as follows:

\[
\mu_A(x; a, b, c, d) = \begin{cases} 
0, & x < a \text{ and } x > d \\
\frac{x - a}{b - a}, & a \leq x \leq b \\
1, & b \leq x \leq c \\
\frac{d - x}{d - c}, & c \leq x \leq d
\end{cases}
\]

where \( A = [a, b, c, d] \) is called a trapezoidal fuzzy number.

Especially, when \( b = c \) in Definition 2.1, we get a triangular membership function. We give a definition of triangular membership function in the following.
2.1 Fuzzy Numbers

Definition 2.2 Triangular Membership Function

Let \( A = [a, b, c] \) be a triangular fuzzy number, then its membership function is defined as follows:

\[
\mu_A(x; a, b, c) = \begin{cases} 
0, & x < a \text{ and } x > c \\
\frac{x - a}{b - a}, & a \leq x \leq b \\
\frac{c - x}{c - b}, & b \leq x \leq c 
\end{cases}.
\]

Moreover, if \( a = b \) and \( c = d \) in Definition 2.1, we get a interval value. We give a definition of uniform membership function in the following.

Definition 2.3 Uniform Membership Function

Let \( A = [a, b] \) be an interval value, then its membership function is defined as follows:

\[
\mu_A(x; a, b) = \begin{cases} 
1, & a \leq x \leq b \\
0, & \text{otherwise}
\end{cases}.
\]

Zadeh (83) proposed fuzzy set theory to deal with the vagueness in data, where membership grade of a fuzzy set is a value between 0 and 1. The following definitions of fuzzy numbers will be used in the whole thesis.

Definition 2.4 (50) Let \( U \) be a universal set and \( C = \{C_1, C_2, \ldots, C_n\} \) be the subset of a specified collection of elements in \( U \). For any term or statement \( X \) on \( U \), the membership function of \( \{C_1, C_2, \ldots, C_n\} \) is denoted \( \{\mu_1(X), \mu_2(X), \ldots, \mu_n(X)\} \), where \( \mu : U \rightarrow [0, 1] \) is a real value function. If the domain of the universal set is discrete, then the fuzzy number \( x \) of \( X \) can be written as

\[
\mu_U(X) = \sum_{i=1}^{n} \mu_i(X) I_{C_i}(X),
\]

(2.1)

where \( I_{C_i}(X) = 1 \) if \( x \in C_i \), and \( I_{C_i}(X) = 0 \) if \( x \notin C_i \).

If the domain of the universal set is continuous, then the fuzzy number \( x \) can be written as

\[
\mu_U(X) = \int_{\bigcap_{C_i \subseteq C} X} \mu_i(X) I_{C_i}(X) dC.
\]

(2.2)

Note that, many writings denote a fuzzy number as

\[
\mu_U(X) = \frac{\mu_1(X)}{C_1} + \frac{\mu_2(X)}{C_2} + \cdots + \frac{\mu_n(X)}{C_n},
\]
where "+" stands for "or", and ":" denotes the membership $\mu_i(X)$ on $C_i$.

### 2.2 Fuzzy Statistic Analysis

**Definition 2.5 Fuzzy Sample Mean (data with interval values)** (50)

Let $U$ be a universe set and \{ $F_i = [a_i, b_i], a_i, b_i \in \mathcal{R}, i = 1, \cdots, n$ \} be a sequence of a random fuzzy sample on $U$. Then, the fuzzy sample mean value is defined as

$$\bar{F} = \left[ \frac{1}{n} \sum_{i=1}^{n} a_i, \frac{1}{n} \sum_{i=1}^{n} b_i \right].$$

**Example 2.1** Let $F_1 = [2, 3]$, $F_2 = [3, 4]$, $F_3 = [4, 6]$, $F_4 = [5, 8]$, and $F_5 = [3, 7]$ be the starting salary for 5 newly graduated master’s students. Then, the fuzzy sample mean for the starting salary of the graduated students will be

$$\bar{F} = \left[ \frac{2 + 3 + 4 + 5 + 3}{5}, \frac{3 + 4 + 6 + 8 + 7}{5} \right] = [3.4, 5.6].$$

### 2.3 Nonparametric Statistical Method

In this section, we will introduce a conventional statistical test which we will use in the following chapters.

#### 2.3.1 Kolmogorov-Smirnov Two-Sample Test

The Kolmogorov-Smirnov Two-Sample Test (hereafter, K-S two-sample test) is designed to evaluate whether two independent samples have been drawn from the same population (or from populations with the same distribution). A two-tailed test is sensitive to any kind of difference in the distributions from which the two samples are drawn. A one-tailed test is used to decide whether the sample values in the population of samples are stochastically larger than the values of the population of the other samples.

To apply the K-S two-sample test, we determine the cumulative frequency distribution for each sample of observations by using the same intervals for both distributions. Then, for each interval, we subtract one step function from the other. The test focuses on the largest one of these observed deviations.
2.3 Nonparametric Statistical Method

Let \( S_m(X) \) be the empirical distribution function for one sample of size \( m \), that is, 
\[
S_m(X) = \frac{1}{m} \sum_{i=1}^{m} I_{X_i \leq x},
\]
where \( I_{X_i \leq x} \) is the indicator function, equal to 1 if \( X_i \leq x \) and equal to 0 otherwise. Let \( S_n(X) \) be the empirical distribution function for the other sample of size \( n \), that is, 
\[
S_n(X) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq x}.
\]
Thus, the K-S two-sample test statistic is
\[
D_{m,n} = \sup \{ S_m(X) - S_n(X) \},
\]
for a one-tailed test, and it is
\[
D_{m,n} = \sup \{ |S_m(X) - S_n(X)| \},
\]
for a two-tailed test. Note that equation (2.4) uses the absolute value.

In each case, the sampling distribution of \( D_{m,n} \) is known. The probabilities associated with observed values as large as the observed \( D_{m,n} \) under the null hypothesis (i.e., the two samples have come from the same distribution) are tabled in Reference (47). In fact, there may be two sampling distributions, depending upon whether the test is one-tailed or two-tailed. Notice that for a one-tailed test, we observe \( D_{m,n} \) in the predicted direction using equation (2.3), and for a two-tailed test, we observe the maximum absolute difference \( D_{m,n} \) using equation (2.4), regardless of the direction. This is because in the one-tailed test, \( H_1 \) indicates that the population values relating to one of the samples are stochastically larger than the population values relating to the other sample, whereas in the two-tailed test, \( H_1 \) simply indicates that the two samples are from different populations.

If both \( m \) and \( n \) are 25 or less, Appendix Table \( L_I \) in Reference (63) can be used as a reference to test the null hypothesis against a one-tailed alternative, and Appendix Table \( L_{II} \) in Reference (63) can be used as a reference to test the null hypothesis against a two-tailed alternative. These tables give values for \( D_{m,n} \) that are significant at various levels. The critical values of the test statistic can be derived if values of \( m, n, mnD_{m,n} \) as well as whether the tests that are one-tailed are known.

When either \( m \) or \( n \) are larger than 25, Appendix Table \( L_{III} \) in Reference (63) may be used for the K-S two-sample test. To use this table, determine the value of \( D_{m,n} \) for observed data by using the following equation:
\[
K(\alpha) \sqrt{\frac{m + n}{mn}},
\]
(2.5)
2. SOME BASIC THEORETICAL CONCEPTS

where \( \alpha \) is the significant level and the value of coefficient \( K(\alpha) \) can be found in Table L_{III} of Reference (63).

Hence, we present the steps of the K-S two-sample test.

**Step 1.** Arrange both the groups of scores in a cumulative frequency distribution using the same intervals (or classifications) for both distributions. Use as many intervals as possible.

**Step 2.** Using subtraction, determine the difference between the two-sample cumulative distributions.

**Step 3.** Determine the largest of the differences \( D_{m,n} \). For a one-tailed test, \( D_{m,n} \) is the largest difference in the predicted direction. For a two-tailed test, \( D_{m,n} \) is the largest difference in either direction.

**Step 4.** Determine the significance of the observed values \( D_{m,n} \) depending on the sample size and the nature of \( H_1 \). When \( m \) and \( n \) are both \( \leq 25 \), Appendix Table L_{I} in Reference (63) is referenced for the one-tailed test, and Appendix Table L_{III} in Reference (63) is referenced for the two-tailed test. In both tables, entry \( mnD_{m,n} \) is used. For a two-tailed test when either \( m \) or \( n \) are larger than 25, Appendix Table L_{III} in Reference (63) is used. Critical values of \( D_{m,n} \) for any given large values of \( m \) or \( n \) may be computed by using the formula (2.5).

**Step 5.** If the observed value is equal to or greater than that given in the appropriate table for a particular level of significance, \( H_0 \) may be rejected in favor of \( H_1 \).

We show a part of Tables of Reference (63) in our Appendix A which we will use in the chapters 3 and 4.

Similar to the K-S one-sample test, the K-S two-sample test focuses on the agreement between two cumulative distributions. If the two samples have indeed been drawn from the same population distribution, then the cumulative distributions of both the samples should be expected to be fairly close to each other, inasmuch as they both should show only random derivations from the common population distribution. If the two fuzzy sample cumulative distributions are too far apart at any point, this suggests that the samples come from different populations. Thus, a sufficiently large deviation between the two sample cumulative distributions is evidence to reject \( H_0 \).
3

Kolmogorov-Smirnov Two Sample Test with Continuous Fuzzy Data

3.1 Introduction

The Kolmogorov-Smirnov (K-S) two-sample test is a goodness-of-fit test that is used to determine whether two underlying distributions differ. It is customary to call the K-S two-sample test the Smirnov test (64), while the Kolmogorov test is sometimes called the K-S one-sample test. In this chapter, we discuss only the K-S two-sample test, as our purpose here is to test whether two independent samples have been drawn from the same population. The two-sample test is one of the most useful nonparametric methods for comparing two samples, as it is sensitive to differences in both the location and the shape of the empirical cumulative distribution functions of the two samples. Other tests, such as the median test, the Mann-Whitney test, or the parametric *t*-test, may also be appropriate (6). However, while these tests are sensitive to differences between two means or medians, they may not detect other types of differences, such as differences in variance. One of the advantages of two-tailed tests is that such tests consistently reflect all types of differences between two distribution functions. Although many papers have discussed the powerful K-S two-sample test (see the discussions in (9), (10), and (60)), these reports have all simulated these tests based on known distributions. However, vague information is sometimes given when describing data in natural language, and the
underlying distribution of the fuzzy data is not known. It can be difficult to put such information into statistical terms; therefore, we must establish techniques to handle such information.

In this chapter, we propose a method of judging whether two continuous fuzzy data samples have been drawn from the same population. We use the K-S two-sample test to address this problem. However, the K-S two-sample test is concerned with real numbers. To manipulate continuous fuzzy data by means of the K-S two-sample test, we must find a method for classifying all of the continuous fuzzy data. Accordingly, we propose some new rules for classifying and ranking continuous fuzzy data. Several ranking methods have previously been proposed for fuzzy numbers; for instance, Chen (8) used the distance between the fuzzy numbers and the comparison data to find the greatest distance. Similar to Kaufmann and Gupta (22), Liou and Wang (41) use a membership function to rank fuzzy numbers. Yager (82) proposed a method of ranking fuzzy numbers using a centroid index. Although there are many ways to rank fuzzy numbers, all the methods that have been used are based on the central point. Any such method will lose some information about continuous fuzzy data. Thus, we use a weight function to rank fuzzy numbers. The weight function includes both the central point and the radius, which can be used to classify all continuous fuzzy data. When we use this information, the K-S two-sample test with continuous fuzzy data can be implemented.

3.2 Preliminary Preparation

First, we give some definitions we will use in this chapter. We determine the fuzzy data as central point and radius by using following definition.

**Definition 3.1 Moments and Center of Mass of a Planar Lamina (33)**

Let $f$ and $g$ be continuous functions such that $f(x) \geq g(x)$ on $[a,b]$, and consider the planar lamina of uniform density $\rho$ bounded by the graphs of $y = f(x)$, $y = g(x)$, and $a \leq x \leq b$.

1. The moments about the $x$–axis and $y$–axis are

$$M_x = \rho \int_a^b \frac{(f(x) + g(x))}{2} |f(x) - g(x)| \, dx$$  \hspace{1cm} (3.1)
3.3 Kolmogorov-Smirnov Two-Sample Test with Continuous Fuzzy Data

\[ M_y = \rho \int_a^b x[f(x) - g(x)]dx. \quad (3.2) \]

2. The center of mass \((\overline{x}, \overline{y})\) is given by \(\overline{x} = \frac{M_y}{m}\) and \(\overline{y} = \frac{M_x}{m}\), where \(m = \rho \int_a^b [f(x) - g(x)]dx\) is the mass of the lamina.

Note that in mathematics, a planar lamina is a closed surface of mass \(m\) and surface density \(\rho\). It can be used to determine moments of inertia, or center of mass.

We also need an definition as follows. We will use it to define our function in this thesis.

**Definition 3.2 The Mean Value Theorem for Definite Integrals** (14)

If \(f\) is continuous on \([a, b]\), then there exists some points \(c\) in \([a, b]\), such that

\[ f(c) = \frac{1}{b-a} \int_a^b f(x)dx. \]

3.3 Kolmogorov-Smirnov Two-Sample Test with Continuous Fuzzy Data

3.3.1 Empirical distribution function with continuous fuzzy data

In order to provide the empirical distribution function for continuous fuzzy data, we must classify the continuous fuzzy data. We first define a weight function for continuous fuzzy data, and then use it to pursue a new classification. Thus, the empirical distribution function for the continuous fuzzy data can be found.

In order to correct the data accurately, we use the continuous revising to define the weight function as follows.

**Definition 3.3 Weight function for continuous fuzzy data**

The weight function of continuous fuzzy data \(X_i \equiv (o_i, l_i)\) is defined as follows:

\[ WF_{x_i} \equiv WF(o_i, l_i) = o_i[1 + ke^{-2l_i}], \forall i = 1, 2, 3, \ldots \quad (3.3) \]

where \(o_i\) is the central point, \(l_i\) is the radius with respect to \(o_i\), and \(k = \max(o_i + l_i) - \min(o_j - l_j), \forall i, j = 1, 2, 3, \ldots \). We name \(k\) as weight constant.
3. KOLMOGOROV-SMIRNOV TWO SAMPLE TEST WITH CONTINUOUS FUZZY DATA

Property 3.1 Let $X_i = [a_i, b_i]$ be an interval value, then $o_i = \frac{a_i + b_i}{2}, l_i = \frac{b_i - a_i}{2}$, and $k = \max b_i - \min a_j, \forall i, j = 1, 2, 3, \ldots$

Proof: It is trivial that $o_i = \frac{a_i + b_i}{2}$ and $l_i = \frac{b_i - a_i}{2}$.

Therefore, we have

$$k = \max (o_i + l_i) - \min (o_j - l_j)$$

$$= \max (\frac{a_i + b_i}{2} + \frac{b_i - a_i}{2}) - \min (\frac{a_j + b_j}{2} - \frac{b_j - a_j}{2})$$

$$= \max b_i - \min a_j, \forall i, j = 1, 2, 3, \ldots$$

Property 3.2 Let $X_i = [a_i, b_i, c_i]$ be triangular fuzzy numbers, then $o_i = \frac{a_i + b_i + c_i}{3}, l_i = \frac{c_i - a_i}{4},$ and $k = \max (\frac{a_i + 4b_i + 7c_i}{12}) - \min (\frac{7a_i + 4b_i + c_i}{12}), \forall i, j = 1, 2, 3, \ldots$

Proof: By Definition 3.1, we let $\rho = 1$ and we can find that $o_i = \frac{M_y}{m}$.

When $X_i$ is a triangular fuzzy number, its membership function is denoted as follows:

$$f(x) = \begin{cases} 0, & x < a \text{ and } x > c \\ \frac{x - a}{b - a}, & a \leq x \leq b \\ \frac{c - x}{c - b}, & b \leq x \leq c \end{cases}.$$ 

Therefore, $M_y = 1 \int_a^b \frac{x - a}{b - a} dx + 1 \int_b^c \frac{c - x}{c - b} dx = \frac{1}{6} (c - a) (a + b + c)$ and $m = 1 \int_a^b \frac{x - a}{b - a} dx + 1 \int_b^c \frac{c - x}{c - b} dx = \frac{c - a}{2}$.

Hence,

$$o_i = \frac{M_y}{m} = \frac{1}{6} \frac{(c_i - a_i)(a_i + b_i + c_i)}{\frac{c_i - a_i}{2}} = \frac{a_i + b_i + c_i}{3}.$$ 

Moreover, from Definition 3.2, the mean value theorem for definite integrals (14) enables us to find some points $t$ in $[a, c]$ such that

$$(c - a)f(t) = \int_a^c f(x)dx = \frac{c - a}{2}.$$ 

Therefore, $f(t) = \frac{1}{2}, \forall t \in [a, c]$.

In the case where there are two points, say $t_1$ and $t_2$, such that

$$f(t_1) = f(t_2) = \frac{1}{2}, \forall t_1, t_2 \in [a, c].$$ 

16
3.3 Kolmogorov-Smirnov Two-Sample Test with Continuous Fuzzy Data

This results in \( t_1 = \frac{a + b}{2} \) and \( t_2 = \frac{b + c}{2} \).

There also exists a rectangle with the same area as \( \frac{c - a}{2} \). (See Figure 3.1.)

Hence \( 2l = t_2 - t_1 = \frac{c - a}{2}, l = \frac{c - a}{4} \).

When we have \( o_i \) and \( l_i \), the weight constant \( k \) is

\[
k = \max(o_i + l_i) - \min(o_j - l_j)
\]

\[
= \max\left(\frac{a_i + b_i + c_i - a_i}{3} + \frac{c_i - a_i}{4}\right) - \min\left(\frac{a_i + b_i + c_i}{3} - \frac{c_i - a_i}{4}\right)
\]

\[
= \max\left(\frac{a_i + 4b_i + 7c_i}{12}\right) - \min\left(\frac{7a_j + 4b_j + c_i}{12}\right), \forall i, j = 1, 2, 3, \ldots
\]

**Property 3.3** Let \( X_i = [a_i, b_i, c_i, d_i] \) be trapezoidal fuzzy numbers, then

\[
o_i = \frac{(c_i + d_i)^2 - (a_i + b_i)^2 + a_ib_i - c_id_i}{3[(c_i + d_i) - (a_i + b_i)]}, \quad l_i = \frac{(c_i + d_i) - (a_i + b_i)}{4}, \quad \text{and } k = \max(o_i + l_i) - \min(o_j - l_j), \forall i, j = 1, 2, 3, \ldots
\]

**Proof:** By Definition 3.1, we let \( \rho = 1 \) and we can find that \( o_i = \frac{M_y}{m} \).

When \( X_i \) is a trapezoidal fuzzy number, its membership function is denoted as
3. KOLMOGOROV-SMIRNOV TWO SAMPLE TEST WITH CONTINUOUS FUZZY DATA

follows:

\[ f(x) = \begin{cases} 
0, & x < a \\
\frac{x-a}{b-a}, & a \leq x \leq b \\
1, & b \leq x \leq c \\
\frac{d-x}{d-c}, & c \leq x \leq d 
\end{cases} \]

Therefore, \( M_y = 1 \int_a^b \frac{x-a}{b-a} dx + 1 \int_b^c x dx + 1 \int_c^d \frac{d-x}{d-c} dx = \frac{1}{6}[(c+a)^2 - (a+b)^2 + (ab-cd)] \) and \( m = 1 \int_a^b \frac{x-a}{b-a} dx + 1 \int_b^c x dx + 1 \int_c^d \frac{d-x}{d-c} dx = \frac{1}{2}[(c+d) - (a+b)]. \)

Hence,

\[ o_i = \frac{M_y}{m} = \frac{\frac{1}{6}[(c+a)^2 - (a+b)^2 + (ab-cd)]}{\frac{1}{2}[(c+d) - (a+b)]} = \frac{[(c+d)^2 - (a+b)^2 + (ab-cd)]}{3[(c+d) - (a+b)]}. \]

Moreover, from Definition 3.2, the mean value theorem for definite integrals (14) enables us to find some points \( t \) in \([a,d]\) such that

\[ (d-a)f(t) = \int_a^d f(x)dx = \frac{(c+d) - (a+b)}{2}. \]

Therefore, \( f(t) = \frac{(c+d)(a+b)}{2(d-a)}, \forall t \in [a,d]. \)

In the case where there are two points, say \( t_1 \) and \( t_2 \), such that

\[ f(t_1) = f(t_2) = \frac{(c+d)(a+b)}{2(d-a)}, \forall t_1, t_2 \in [a,d]. \]

We can also find a rectangle with the same area as \( \frac{(c+d)(a+b)}{2}. \)

Hence, \( 2l = t_2 - t_1 = \frac{(c+d)(a+b)}{2} \) and \( l = \frac{(c+d)(a+b)}{4}. \)

When we have \( o_i \) and \( l_i \), the weight constant \( k \) is

\[ k = \max(o_i + l_i) - \min(o_j - l_j), \forall i, j = 1, 2, 3 \ldots \]

Definition 3.4 Fuzzy classification

If \( WF_{x_i} < WF_{x_j}, \forall i \neq j \), we say that \( x_i \) and \( x_j \) are in different classes. In particular, \( x_i \) is the class before \( x_j \). Moreover, if \( WF_{x_i} = WF_{x_j}, \forall i \neq j \), we say that \( x_i \) and \( x_j \) are in the same class.

Definition 3.5 Identical independence of continuous fuzzy data

If \( WF_{x_i} \neq WF_{x_j}, \forall i \neq j \), we say that \( x_i \) and \( x_j \) are identical independent by the choice of \( k \) (weight constant). Otherwise, \( x_i \) and \( x_j \) are dependent.
3.3 Kolmogorov-Smirnov Two-Sample Test with Continuous Fuzzy Data

Definition 3.6 Empirical distribution function with continuous fuzzy data

Let \( x_1, x_2, \ldots, x_n \) be \( n \) continuous fuzzy data. We can use the weight function (WF) to separate \( x_i \) into different class \( c_i \), which are called Glivenko-Cantelli classes (see discussion in References (13), (15), and (61)). If \( x_i \) and \( x_j \) are in different classes, then we say that \( x_i \) and \( x_j \) are identically independent for \( i \neq j \). Moreover, we have the order statistic of \( x_i \) (assume that they are in different classes), denoted as

\[
WF_{x(1)} < WF_{x(2)} < \cdots < WF_{x(n)}
\]  

(3.4)

Hence, the empirical distribution function can be generalized to a set \( C \) to obtain an empirical measure indexed by \( c_i \).

\[
S_n(c_i) = \frac{1}{n} \sum_{i=1}^{n} I_{c_i}(WF_{x_i}), c_i \in C,
\]

(3.5)

where \( I_{c_i} \) is the indicator function denoted by

\[
I_{c_i}(WF_{x_i}) = \begin{cases} 
1, & WF_{x_i} \in c_i, \\
0, & WF_{x_i} \notin c_i, \forall i = 1, 2, \ldots n.
\end{cases}
\]  

(3.6)

Now, when we have those definitions, we can proceed to study the Kolmogorov-Smirnov two-sample test with continuous fuzzy data.

3.3.2 Kolmogorov-Smirnov two-sample test with continuous fuzzy data

Procedure for using K-S two-sample test for continuous fuzzy data (Two-tailed test) in small samples:

1. Samples: Let \( X_m \) and \( Y_n \) be two samples with continuous fuzzy data. \( X_i \) has size \( m \) and \( Y_j \) has size \( n \). Combining all observations, we have \( N = m + n \) pieces of data. A value of the weight function \( WF \) can be found that will let us distribute \( X_m \) and \( Y_n \) into different classes \( c_i \) (maybe in the same class). The number of classes is less than or equal to \( N \). Moreover, the two empirical distribution functions of \( X_m \) and \( Y_n \) can be found individually.

2. Hypothesis: Two samples have the same distribution \( H_0 \).

3. Statistics: \( D_{m,n} = max|S_m(X) - S_n(X)| \).

4. Decision rule: Under significance level \( \alpha \). Appendix A Table I is used.
3.4 Empirical Studies

Example 3.1 A Japanese dining hall manager planned to introduce new boxed lunch services and decided to take a survey to investigate what price for a boxed lunch would be acceptable to male and female customers. A sample was randomly selected of 20 customers (10 males and 10 females) who resided around this dining hall in the city of Taipei. The investigator asked them, how many dollars they would be willing to spend (can answer with interval) for a boxed lunch in a Japanese dining hall. The answers are shown in Table 3.1.

Table 3.1: The Price which will be Acceptable by Males and Females

<table>
<thead>
<tr>
<th>Males</th>
<th>[60,70]</th>
<th>[70,90]</th>
<th>[50,80]</th>
<th>[60,100]</th>
<th>[50,80]</th>
<th>[50,70]</th>
<th>[50,100]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>[50,60]</td>
<td>[60,70]</td>
<td>[80,100]</td>
<td>[90,120]</td>
<td>[90,100]</td>
<td>[55,75]</td>
<td>[70,90]</td>
</tr>
<tr>
<td></td>
<td>[100,120]</td>
<td>[80,120]</td>
<td>[90,120]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

First, we distributed male answers and female answers into different classes. We had to find the weight values and compare them. Moreover, we had to determine which class they belong to. The calculation was done as Table 3.2.

Comparison among $WF$, results in the following inequality:

$WF_{X_4} = WF_{Y_1} < WF_{X_8} < WF_{X_3} = WF_{X_7} < WF_{Y_6} < WF_{X_1} = WF_{Y_2} < WF_{X_{10}} < WF_{X_9} < WF_{X_2} = WF_{X_6} = WF_{Y_7} < WF_{X_5} = WF_{Y_5} < WF_{Y_3} < WF_{Y_4} = WF_{Y_{10}} < WF_{Y_8}$.

Here, we take $k = max\ b_i - min\ a_j = 120 - 50 = 70, \forall i, j = 1, 2, \ldots, 20$.

From the above, we have 13 classes. Now, we went on to find the cumulative distributions of $X_i$ and $Y_j$.

From Table 3.3, the test statistic was obtained:

$$D = max|S_{10}(X) - S_{10}(Y)| = 0.5.$$ 

at a significance level $\alpha = 0.05$, $mnD = 10 \times 10 \times (0.5) = 50 < 70$ (Appendix A Table I). Since the observed value did not exceed the critical value, we did not reject $H_0$. We conclude that males and females have the same interval of the acceptable price of a boxed lunch.

We use the same data in Table 3.1 and different method to defuzzify the interval values. We denote this method as $C$ method which is a conventional method by using
### 3.4 Empirical Studies

#### Table 3.2: The Weight Values and Classes

<table>
<thead>
<tr>
<th>$[a_i, b_i]$</th>
<th>$a_i$</th>
<th>$b_i$</th>
<th>$WF$</th>
<th>$WF'$ of $WF$</th>
<th>$C_i$ of $WF$</th>
<th>$C_i$ of $a_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ [60, 70]</td>
<td>65 5</td>
<td>65$[1 + ke^{-10}]$</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_2$ [70, 90]</td>
<td>80 10</td>
<td>80$[1 + ke^{-20}]$</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_3$ [50, 80]</td>
<td>65 15</td>
<td>65$[1 + ke^{-30}]$</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_4$ [50, 60]</td>
<td>55 5</td>
<td>55$[1 + ke^{-10}]$</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_5$ [80,100]</td>
<td>90 10</td>
<td>90$[1 + ke^{-20}]$</td>
<td>9</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_6$ [70, 90]</td>
<td>80 10</td>
<td>80$[1 + ke^{-20}]$</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_7$ [50, 80]</td>
<td>65 15</td>
<td>65$[1 + ke^{-30}]$</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_8$ [50, 70]</td>
<td>60 10</td>
<td>60$[1 + ke^{-20}]$</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_9$ [65, 95]</td>
<td>80 15</td>
<td>80$[1 + ke^{-30}]$</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_{10}$ [50,100]</td>
<td>75 25</td>
<td>75$[1 + ke^{-50}]$</td>
<td>6</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_1$ [50, 60]</td>
<td>55 5</td>
<td>55$[1 + ke^{-10}]$</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_2$ [60, 70]</td>
<td>65 5</td>
<td>65$[1 + ke^{-10}]$</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_3$ [80,100]</td>
<td>90 10</td>
<td>90$[1 + ke^{-20}]$</td>
<td>9</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_4$ [90,120]</td>
<td>105 15</td>
<td>105$[1 + ke^{-30}]$</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_5$ [90,100]</td>
<td>95 5</td>
<td>95$[1 + ke^{-10}]$</td>
<td>10</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_6$ [55, 75]</td>
<td>65 15</td>
<td>65$[1 + ke^{-30}]$</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_7$ [70, 90]</td>
<td>80 15</td>
<td>80$[1 + ke^{-30}]$</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_8$ [100,120]</td>
<td>110 15</td>
<td>110$[1 + ke^{-30}]$</td>
<td>13</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_9$ [80,120]</td>
<td>100 20</td>
<td>100$[1 + ke^{-40}]$</td>
<td>11</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{10}$ [90,120]</td>
<td>105 15</td>
<td>105$[1 + ke^{-30}]$</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Table 3.3: The Cumulative Distributions of $X_i$ and $Y_j$

<table>
<thead>
<tr>
<th>$C_i$ of $WF$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{10}(X)$</td>
<td>.1</td>
<td>.2</td>
<td>.4</td>
<td>.4</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$S_{10}(Y)$</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.9</td>
<td>1</td>
</tr>
<tr>
<td>$</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>$</td>
<td>0</td>
<td>.1</td>
<td>.3</td>
<td>.2</td>
<td>.2</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.4</td>
<td>.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$C_i$ of $a_i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{10}(X)$</td>
<td>.1</td>
<td>.2</td>
<td>.5</td>
<td>.6</td>
<td>.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$S_{10}(Y)$</td>
<td>.1</td>
<td>.1</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.9</td>
<td>1</td>
</tr>
<tr>
<td>$</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>$</td>
<td>0</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.4</td>
<td>.4</td>
<td>.5</td>
<td>.6</td>
</tr>
</tbody>
</table>
central point $o_i$ to defuzzify the fuzzy data. The classes of $C$ method is shown in Table 3.2. We also give the cumulative distributions of $X_i$ and $Y_j$ in Table 3.3.

From Table 3.3, the test statistic of $C$ method was obtained:

$$D = \max |S_{10}(X) - S_{10}(Y)| = 0.8.$$ 

at a significance level $\alpha = 0.05$, $mnD = 10 \times 10 \times (0.5) = 80 > 70$ (Appendix A Table I). Since the observed value exceeds the critical value, we reject $H_0$. We conclude that males and females have different interval of the acceptable price of a boxed lunch. 

**Example 3.2** With the rest of the procedure as mentioned in Example 3.1, the investigator asked the 20 customers in the following questions: 1. In which price (interval values) they would be willing to spend for a lunch box in a Japanese dining hall? 2. In which price (real numbers) they would most possibility buy the lunch box in a Japanese dining hall? We can collect those data and get triangular fuzzy numbers. The answers are shown in Table 3.4.

| Table 3.4: The Price which will be Acceptable by Males and Females |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Males             | [ 0, 60,100]      | [ 60, 70,100]     | [ 30, 60,100]     | [ 50, 65, 80]     | [ 50, 60,100]     |
|                   | [ 50, 50, 80]     | [ 60, 65, 80]     | [ 60, 60, 80]     | [ 50, 80,160]     | 40,100,150        |
| Females           | [ 40, 60, 70]     | 50, 60,100       | [ 50, 60,100]     | [150,150,300]     | 50, 60, 70        |
|                   | [ 50, 70, 80]     | [ 40, 60,150]    | [ 50, 60, 80]     | [ 50,100,200]     | 50, 70,200        |

First, we distributed male answers and female answers into different classes. We had to find the weight values and compare them. Moreover, we had to determine which class they belong to. The calculation was done as Table 3.5.

Comparison among $WF$, results in the following inequality:

$$WF_{X_1} < WF_{Y_1} < WF_{X_6} < WF_{Y_5} < WF_{X_3} < WF_{Y_8} < WF_{X_4} < WF_{X_9} = WF_{Y_6} < WF_{X_7} < WF_{X_5} = WF_{Y_2} = WF_{Y_3} < WF_{X_2} < WF_{Y_7} < WF_{X_9} = WF_{X_{10}} < WF_{Y_{10}} < WF_{Y_9} < WF_{Y_4}$$

Here, we take $k = \max (o_i + l_i) - \min (o_j - l_j) = 237.5 - \frac{85}{1} \approx 209.17, \forall i, j = 1, 2, \ldots, 20$.

From the above, we have 16 classes. Next, we went on to find the cumulative distributions of $X_i$ and $X_j$. 

22
3.4 Empirical Studies

Table 3.5: The Weight Values and Classes

<table>
<thead>
<tr>
<th></th>
<th>$[a_i,b_i]$</th>
<th>$o_i$</th>
<th>$l_i$</th>
<th>$WF$</th>
<th>$\mathcal{C}_i$ of $WF$</th>
<th>$\mathcal{C}_i$ of $o_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>[0, 60,100]</td>
<td>53.33</td>
<td>25.00</td>
<td>53.33$[1 + ke^{-50}]$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$X_2$</td>
<td>[60, 70,100]</td>
<td>76.67</td>
<td>10.00</td>
<td>76.67$[1 + ke^{-20}]$</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>$X_3$</td>
<td>[30, 60,100]</td>
<td>63.33</td>
<td>17.50</td>
<td>63.33$[1 + ke^{-35}]$</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>$X_4$</td>
<td>[50, 65, 80]</td>
<td>65.00</td>
<td>7.50</td>
<td>65.00$[1 + ke^{-15}]$</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>$X_5$</td>
<td>[50, 60,100]</td>
<td>70.00</td>
<td>12.50</td>
<td>70.00$[1 + ke^{-25}]$</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>$X_6$</td>
<td>[50, 50, 80]</td>
<td>60.00</td>
<td>7.50</td>
<td>60.00$[1 + ke^{-15}]$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$X_7$</td>
<td>[60, 65, 80]</td>
<td>68.33</td>
<td>5.00</td>
<td>68.33$[1 + ke^{-10}]$</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>$X_8$</td>
<td>[60, 60, 80]</td>
<td>66.67</td>
<td>7.50</td>
<td>66.67$[1 + ke^{-15}]$</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>$X_9$</td>
<td>[50, 80,100]</td>
<td>96.67</td>
<td>27.50</td>
<td>96.67$[1 + ke^{-55}]$</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>[40,100,150]</td>
<td>96.67</td>
<td>27.50</td>
<td>96.67$[1 + ke^{-55}]$</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>$Y_1$</td>
<td>[40, 60, 70]</td>
<td>56.67</td>
<td>7.50</td>
<td>56.67$[1 + ke^{-15}]$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>[50, 60,100]</td>
<td>70.00</td>
<td>12.50</td>
<td>70.00$[1 + ke^{-25}]$</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>[50, 60,100]</td>
<td>70.00</td>
<td>12.50</td>
<td>70.00$[1 + ke^{-25}]$</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>$Y_4$</td>
<td>[150,150,300]</td>
<td>200.00</td>
<td>37.50</td>
<td>200.00$[1 + ke^{-75}]$</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>$Y_5$</td>
<td>[50, 60, 70]</td>
<td>60.00</td>
<td>5.00</td>
<td>60.00$[1 + ke^{-10}]$</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>$Y_6$</td>
<td>[50, 70, 80]</td>
<td>66.67</td>
<td>7.50</td>
<td>66.67$[1 + ke^{-15}]$</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>$Y_7$</td>
<td>[40, 60,150]</td>
<td>83.33</td>
<td>27.50</td>
<td>83.33$[1 + ke^{-55}]$</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>$Y_8$</td>
<td>[50, 60, 80]</td>
<td>63.33</td>
<td>7.50</td>
<td>63.33$[1 + ke^{-15}]$</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>$Y_9$</td>
<td>[50,100,200]</td>
<td>116.67</td>
<td>37.50</td>
<td>116.67$[1 + ke^{-75}]$</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>$Y_{10}$</td>
<td>[50, 70,200]</td>
<td>106.67</td>
<td>37.50</td>
<td>106.67$[1 + ke^{-75}]$</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 3.6: The Cumulative Distributions of $X_i$ and $X_j$

<table>
<thead>
<tr>
<th>$\mathcal{C}_i$ of $WF$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{10}(X)$</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.8</td>
<td>.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$S_{10}(Y)$</td>
<td>0</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.4</td>
<td>.6</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
</tr>
<tr>
<td>$</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>$</td>
<td>.1</td>
<td>0</td>
<td>.1</td>
<td>0</td>
<td>.1</td>
<td>0</td>
<td>.1</td>
<td>1</td>
<td>2</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.2</td>
<td>.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mathcal{C}_i$ of $o_i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{10}(X)$</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.8</td>
<td>.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$S_{10}(Y)$</td>
<td>0</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.4</td>
<td>.6</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
</tr>
<tr>
<td>$</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>$</td>
<td>.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.3</td>
<td>.2</td>
</tr>
</tbody>
</table>
From Table 3.6, the test statistic was obtained:

\[
D = \max\{|S_{10}(X) - S_{10}(Y)| = 0.3.
\]

at a significance level \( \alpha = 0.05 \), \( mnD = 10 \times 10 \times (0.3) = 30 < 70 \) (Appendix A Table I). Since the observed value did not exceed the critical value, we did not reject \( H_0 \). We conclude that males and females have the same interval of the acceptable price of a boxed lunch.

We use the same data in Table 3.4 and \( C \) method to defuzzify the triangular fuzzy numbers. The classes of \( C \) method is shown in Table 3.5 and the cumulative distributions of \( X_i \) and \( Y_j \) is given in Table 3.6.

From Table 3.6, the test statistic of \( C \) method was obtained:

\[
D = \max\{|S_{10}(X) - S_{10}(Y)| = 0.3.
\]

at a significance level \( \alpha = 0.05 \), \( mnD = 10 \times 10 \times (0.3) = 30 < 70 \) (Appendix A Table I). We have the same result as our method in this example.

**Example 3.3** With the rest of the procedure as illustrated in Example 3.1. The investigator asked them in the following questions: 1. When the price is about \( \) (interval values) Taiwan dollars, you would be willing to spend for a lunch box in a Japanese dining hall. 2. When the price is more than \( \) (real numbers) Taiwan dollars, you would not buy the lunch box in a Japanese dining hall. 3. When the price is less than \( \) (real numbers) Taiwan dollars, you would not buy the lunch box in a Japanese dining hall. We can collect those data and get trapezoidal fuzzy numbers. The answers are shown in Table 3.7.

| Males | [0, 60, 90, 100] | [60, 60, 90, 100] | [30, 60, 90, 100] | [50, 60, 80, 80] | [50, 50, 80, 100] |
|-------|----------------|----------------|----------------|----------------|----------------| | Females | [40, 50, 70, 70] | [50, 50, 70, 100] | [50, 50, 100,100] | [150,150,250,300] | [50, 50, 70, 70] | | [50, 70, 80, 80] | [40, 40, 90,150] | [50, 50, 70, 80] | [50, 60,150,200] | [50, 70,150,200] |

First, we classified male answers and female answers into different classes. We had to find the weight values and compare them. Moreover, we had to determine which class they belong to. The calculation was done as shown in Table 3.8.
### Table 3.8: The Weight Values and Classes

<table>
<thead>
<tr>
<th>$[a_i,b_i]$</th>
<th>$o_i$</th>
<th>$l_i$</th>
<th>$WF$</th>
<th>$C_i$ of $WF$</th>
<th>$C_i$ of $o_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ [0, 60, 90, 100]</td>
<td>60.26</td>
<td>32.50</td>
<td>60.26$[1 + k e^{-65.00}]$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$X_2$ [60, 60, 90, 100]</td>
<td>77.62</td>
<td>17.50</td>
<td>77.62$[1 + k e^{-35.00}]$</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>$X_3$ [30, 60, 90, 100]</td>
<td>69.33</td>
<td>25.00</td>
<td>69.33$[1 + k e^{-50.00}]$</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>$X_4$ [50, 60, 80, 80]</td>
<td>67.33</td>
<td>12.50</td>
<td>67.33$[1 + k e^{-25.00}]$</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>$X_5$ [50, 50, 80, 100]</td>
<td>70.42</td>
<td>20.00</td>
<td>70.42$[1 + k e^{-40.00}]$</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>$X_6$ [50, 50, 80, 80]</td>
<td>65.00</td>
<td>15.00</td>
<td>65.00$[1 + k e^{-30.00}]$</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>$X_7$ [55, 65, 75, 80]</td>
<td>68.57</td>
<td>8.75</td>
<td>68.57$[1 + k e^{-17.50}]$</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>$X_8$ [50, 60, 80, 80]</td>
<td>67.33</td>
<td>12.50</td>
<td>67.33$[1 + k e^{-25.00}]$</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>$X_9$ [50, 70,160,160]</td>
<td>104.00</td>
<td>50.00</td>
<td>104.00$[1 + k e^{-100.00}]$</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>$X_{10}$ [40, 60,120,150]</td>
<td>92.75</td>
<td>42.50</td>
<td>92.75$[1 + k e^{-85.00}]$</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>$Y_1$ [40, 50, 70, 70]</td>
<td>57.33</td>
<td>12.50</td>
<td>57.33$[1 + k e^{-25.00}]$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$Y_2$ [50, 50, 70,100]</td>
<td>68.57</td>
<td>17.60</td>
<td>68.57$[1 + k e^{-35.00}]$</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>$Y_3$ [50, 50,100,100]</td>
<td>75.00</td>
<td>25.00</td>
<td>75.00$[1 + k e^{-50.00}]$</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>$Y_4$ [150,150,250,300]</td>
<td>213.33</td>
<td>62.50</td>
<td>213.33$[1 + k e^{-125.00}]$</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>$Y_5$ [50, 50, 70, 70]</td>
<td>39.17</td>
<td>10.00</td>
<td>39.17$[1 + k e^{-20.00}]$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$Y_6$ [50, 70, 80, 80]</td>
<td>69.17</td>
<td>10.00</td>
<td>69.17$[1 + k e^{-20.00}]$</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>$Y_7$ [40, 40, 90,150]</td>
<td>81.88</td>
<td>40.00</td>
<td>81.88$[1 + k e^{-80.00}]$</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>$Y_8$ [50, 50, 70, 80]</td>
<td>62.67</td>
<td>12.50</td>
<td>62.67$[1 + k e^{-25.00}]$</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$Y_9$ [50, 60,150,200]</td>
<td>115.83</td>
<td>60.00</td>
<td>115.83$[1 + k e^{-120.00}]$</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>$Y_{10}$ [50, 70,150,200]</td>
<td>118.26</td>
<td>57.50</td>
<td>118.26$[1 + k e^{-115.00}]$</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

### Table 3.9: The Cumulative Distributions of $X_i$ and $Y_j$

<table>
<thead>
<tr>
<th>$C_i$ of $WF$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{10}(X)$</td>
<td>0</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$S_{10}(Y)$</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.3</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
</tr>
<tr>
<td>$</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>$</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$C_i$ of $o_i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{10}(X)$</td>
<td>0</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$S_{10}(Y)$</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>$</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.1</td>
<td>.0</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. KOLMOGOROV-SMIRNOV TWO SAMPLE TEST WITH CONTINUOUS FUZZY DATA

Comparison among $WF$ results in the following inequality:

$$WF_{Y_5} < WF_{Y_1} < WF_{X_1} < WF_{Y_8} < WF_{X_6} < WF_{X_4} = WF_{X_8} < WF_{X_7} < WF_{Y_2} < WF_{X_3} < WF_{Y_6} < WF_{X_5} < WF_{Y_3} < WF_{X_2} < WF_{Y_7} < WF_{X_{10}} < WF_{X_9} < WF_{Y_9} < WF_{Y_{10}} < WF_{Y_4}$$

Here, we take $k = \max(o_i + l_i) - \min(o_j - l_j) = (\frac{640}{3} + 62.5) - (\frac{2350}{39} - 32.5) \approx 248.0769 \ldots, \forall i, j = 1, 2, \ldots, 20.$

From the above, we have 19 classes. Now, we went on to find the cumulative distributions of $X_i$ and $Y_j$.

From Table 3.9, the test statistic was obtained in the following:

$$D = \max|S_{10}(X) - S_{10}(Y)| = 0.3.$$  

At a significance level $\alpha = 0.05$, $mnD = 10 \times 10 \times (0.3) = 30 < 70$ (Appendix A Table I). Since the observed value did not exceed the critical value, we did not reject $H_0$. We conclude that males and females have the same interval of the acceptable price of a boxed lunch.

We use the same data in Table 3.7 and $C$ method to defuzzify the trapezoidal fuzzy numbers. The classes of $C$ method is shown in Table 3.8 and the cumulative distributions of $X_i$ and $Y_j$ is given in Table 3.9.

From Table 3.9, the test statistic of $C$ method was obtained:

$$D = \max|S_{10}(X) - S_{10}(Y)| = 0.3.$$  

at a significance level $\alpha = 0.05$, $mnD = 10 \times 10 \times (0.3) = 30 < 70$ (Appendix A Table I). We have the same result as our method in this example.

3.5 Summary of Chapter 3

In this chapter, we studied the use of the K-S two-sample test with small samples of continuous fuzzy data. To identify the statistical pivot, we have defined a new function, the weight function, which includes both the central point and the radius. The weight function can be used to classify all continuous fuzzy data. In addition, we can divide fuzzy data samples into different classes. With this rule, the cumulative distribution function can be determined, and we can obtain the statistical pivot of the K-S test with continuous fuzzy data. We also give three empirical studies with different types of
fuzzy data and compare with the conventional method (C method). We list the results under different significant level $\alpha$ in Table 3.10.

From Table 3.10, we can see that we got the same statistic values 0.3 when the data are triangular fuzzy number and trapezoidal fuzzy number. When the data is interval value, we obtained 0.5 for the statistic value in our method (WF method) for K-S two-sample test and obtained 0.8 for the statistic value in conventional method (C method) for K-S two-sample test. It means that we need stronger evidence to confirm the hypothesis when we used conventional method to defuzzify the interval data in using K-S two sample test. Hence, we concluded that our method is more extensive to use K-S two-sample test for continuous fuzzy data that can enable us to judge whether or not two independent samples of continuous fuzzy data come from the same population.

However, we still can identify some open problems that require future investigation.

1. In this chapter, we used only small sample in the empirical studies. When the sample is large, is it still suitable to use the weight function with K-S two sample test?

2. We only consider the weight function which we proposed in this chapter in K-S two sample test. We hope that we will consider other defuzzification methods with K-S two sample test in the future.
4

Identifying the Distribution Differences of Fuzzy Data Based on a Nonparametric Statistical Method

4.1 Introduction and Literature Review

The two-sample test is one of the most useful nonparametric methods for comparing two samples because it is sensitive to the differences between empirical cumulative distribution function with regard to both the locations and the shapes of the two samples. Other nonparametric statistical tests may also be useful (6), such as the median test, Mann-Whitney test and parametric $t$-test. The Kolmogorov-Smirnov (K-S) two-sample test is a goodness-of-fit test used to determine whether the two underlying distributions of the samples differ. In this chapter, we concentrate our discussion on the K-S two-sample test because no statistical method can distinguish two populations of continuous fuzzy data based on their respective distribution functions. Hence, we use the K-S two-sample test to decide whether the two independent samples of continuous fuzzy data are derived from the same population. We denote a sample of continuous fuzzy data as a set of data obtained from a single population. Given two different samples of continuous fuzzy data, our goal is to test whether or not they have been drawn from the same population. This method is useful in various applications,
4.1 Introduction and Literature Review

such as industry, engineering, social surveys, and others.

Although many papers have discussed the powerful K-S two-sample test (see the discussion in (9), (10), (47) and (60)), these studies have virtually always simulated the test under known distributions. However, sometimes vague information is obtained, for example, when data are provided in natural language. When using fuzzy data, the underlying distribution is not known. Moreover, it can be difficult to put such information into statistical terms; therefore, we must establish techniques to handle this information.

To manipulate continuous fuzzy data using the K-S two-sample test, we need to first calculate the empirical distribution function of the continuous fuzzy data. Some method is necessary for classifying all of the continuous fuzzy data. Many research works have proposed various ranking methods to classify fuzzy data. For instance, Lee-Kwang and Lee (31) proposed a method that derives rankings by considering the overall possibility distributions of fuzzy numbers and provides users with a method for evaluation. Tseng and Klein (68) designed an algorithm to rank any amount of fuzzy numbers. Ota et al. (53) developed a variable axis method (VAM) to determine the total ordering of fuzzy numbers. Xu and Sasaki (81) proposed a vertex method to calculate the distances between Grey numbers. Lee and You (35) proposed a ranking method that generates possible ranking sequences of given fuzzy numbers. Kang et al. (21) developed a new fuzzy ranking model based on user preferences. Hung et al. (19) provided a novel accuracy function to evaluate interval-valued fuzzy information based on intuition. Yager (82) proposed a method of ranking fuzzy numbers using a centroid index. Although various methods have been proposed to rank fuzzy numbers, all of these methods are based on the concept of a central point. Any of these methods thus ignore some information about continuous fuzzy data in the calculation. Recently, Cheng (8) used the distance between fuzzy numbers to find the largest distance among data points, as they considered defuzzifying to involve the use of two parameters that are calculated from the fuzzy data. We think, however, that it is more effective to analyze original fuzzy data. We will consider this concept and combine it with the concept of fuzzy statistics in this chapter.

An increasing number of research studies have focused on fuzzy statistical analysis, and different applications have been found in various fields. For example, Esogbue and Song (11) proposed a defuzzification method that is rigorously examined and presented
4. IDENTIFYING THE DISTRIBUTION DIFFERENCES OF FUZZY DATA BASED ON A NONPARAMETRIC STATISTICAL METHOD

in an application of the method to the power system stabilization problem, and Chen and Klein (7) proposed an approach using defuzzification methods for the fuzzy MADM. In addition, Wu and Sun (79) presented a set of real-life situations, in which fuzzy techniques can be naturally reformulated in statistical terms. Moreover, Watada et al. (76) built a fuzzy regression model based on fuzzy random data and resorted the data based on some heuristics. These studies have addressed various problems with using defuzzification techniques to choose the central points of fuzzy numbers. Recently, Wu and Chang (77) evaluated the mean and variance values of interval data based on central point and radius data, but they did not consider statistical tests. In applying the concepts of fuzzy data to statistical tests, Lin et al. (40) also defined a weight function in terms of central point and radius values, but they did not give a seriously proof of the weight function, and they also did not present conventional studies for comparison with their method. Hence, in this chapter, we propose to define a defuzzification formula in terms of central point and radius values and to provide a seriously proof to support our concept. We also discuss various empirical studies in this chapter.

The rest of this chapter is organized as follows. Section 4.2 describes the main method. In Section 4.3, some empirical studies show that fuzzy hypothesis testing is useful in soft computations of continuous fuzzy data in the context of social science research. Moreover, a comparison results is provided in Section 4.4. Finally, some conclusions and topics for further studies are emphasized in Section 4.5.

4.2 Identifying the Distribution Difference of Fuzzy Data Based on a Nonparametric Statistical Method

To identify the distribution difference between two populations of fuzzy data, we propose a statistic pivot for K-S two-sample test for two continuous fuzzy data sets. The first step is to define a function that can realize continuous fuzzy data.

4.2.1 Realization of a Continuous Fuzzy Data

To calculate the empirical distribution function for continuous fuzzy data, we must classify continuous fuzzy data. We first give some properties about central point and radius of continuous fuzzy data and define a function for continuous fuzzy data after these properties. Then we use it to derive a new classification.
4.2 Identifying the Distribution Difference of Fuzzy Data Based on a Nonparametric Statistical Method

Note that by Definition 3.1, we know that for trapezoidal fuzzy number:

\[ o = \frac{M_y}{m} = \frac{\rho \int_a^d x[f(x) - g(x)]dx}{\rho \int_a^d [f(x) - g(x)]dx} = \frac{\int_a^d x f(x)dx}{\int_a^d f(x)dx}, \]

where \( g(x) = 0 \). Note that \( m = \int_a^d f(x)dx \) is the area between the membership function \( f(x) \) and the \( x \)-axis.

When we know the values of \( o \), we can define a new membership function based on the central point \( o \). Its membership function is as follows:

\[
h(x) = \begin{cases} 
1, & o - l \leq x \leq o + l \\
0, & \text{otherwise}
\end{cases}
\]

We say that the membership function \( f(x) \) and \( h(x) \) have the same area between \( x \)-axis (see Figure 4.1).

![Membership function](attachment:image.png)

**Figure 4.1:** A Trapezoidal Fuzzy Number \( f(x) \) with Central Point \( o \) and Radius \( l \) Have The Same Area as \( h(x) \)

From Definition 3.2, the mean value theorem for definite integrals (14), we have

\[ m = \int_a^d f(x)dx = \int_{o-l}^{o+l} h(x)dx = 2l. \]

i.e \( l = \frac{\int_a^d f(x)dx}{2} \).
4. IDENTIFYING THE DISTRIBUTION DIFFERENCES OF FUZZY DATA BASED ON A NONPARAMETRIC STATISTICAL METHOD

Now, we give some properties, central point and radius of continuous fuzzy data, in the following.

**Property 4.1** Let \( x = [a, b] \) be an interval value, then its membership function is

\[
f(x) = \begin{cases} 
1, & a \leq x \leq b \\
0, & \text{otherwise}
\end{cases}
\]

Moreover, \( o = \frac{\int_a^b x \, dx}{\int_a^b dx} = \frac{a + b}{2} \) and \( l = \frac{\int_a^b dx}{2} = \frac{b - a}{2} \).

**Property 4.2** Let \( x = [a, b, c] \) be a triangular fuzzy number, then its membership function is

\[
f(x) = \begin{cases} 
0, & x < a \text{ and } x > c \\
\frac{b - a}{b - a}, & a \leq x \leq b \\
\frac{d - x}{c - b}, & b \leq x \leq c
\end{cases}
\]

Moreover,

\[
o = \frac{\int_a^b x \frac{x - a}{b - a} \, dx + \int_b^c x \frac{c - x}{c - b} \, dx}{\int_a^b x \frac{x - a}{b - a} \, dx + \int_b^c x \frac{c - x}{c - b} \, dx} = \frac{1}{6} \frac{(c - a)(a + b + c)}{c - a} = \frac{a + b + c}{3}
\]

and

\[
l = \frac{\int_a^b x \frac{x - a}{b - a} \, dx + \int_b^c x \frac{c - x}{c - b} \, dx}{2} = \frac{c - a}{2} = \frac{c - a}{4}.
\]

**Property 4.3** Let \( x = [a, b, c, d] \) be a trapezoidal fuzzy number, then its membership function is

\[
f(x) = \begin{cases} 
0, & x < a \text{ and } x > d \\
\frac{x - a}{b - a}, & a \leq x \leq b \\
1, & b \leq x \leq c \\
\frac{d - x}{d - c}, & c \leq x \leq d
\end{cases}
\]
4.2 Identifying the Distribution Difference of Fuzzy Data Based on a Nonparametric Statistical Method

Moreover,

\[ o = \frac{\int_a^b x^* x - a\, dx + \int_b^c x\, dx + \int_c^d x^* d - x\, dx}{\int_a^b x - a\, dx + \int_b^c d - x\, dx + \int_c^d d - x\, dx} \]

\[ = \frac{1}{6}[(c + d)^2 - (a + b)^2 + (ab - cd)] \]

\[ = \frac{1}{2}[(c + d) - (a + b)] \]

\[ = \frac{(c + d)^2 - (a + b)^2 + ab - cd}{3[(c + d) - (a + b)]} \quad \text{and} \]

\[ l = \frac{\int_a^b x - a\, dx + \int_b^c x\, dx + \int_c^d d - x\, dx}{\int_a^b b - a\, dx + \int_b^c c - d\, dx + \int_c^d d - c\, dx} \]

\[ = \frac{1}{2}[(c + d) - (a + b)] = \frac{(c + d) - (a + b)}{4}. \]

**Definition 4.1 Realization of a continuous fuzzy data**

Let \( x \equiv (o; l) \) be a continuous fuzzy value on \( U \), which is the university set, \( o = \frac{M_y}{m} \) is the central point and \( l = \frac{m}{2} \) is the radius of fuzzy data. \( M_y \) is the moments about \( y \)-axes and \( m \) is the mass of the lamina in Definition 3.1. The realization of the continuous fuzzy value is defined as follows:

\[ RF_x = o + [1 - e^{-l}], \quad (4.1) \]

which is used to rank the fuzzy data.

It is straight forward that the function is well-defined function because of satisfying the axioms for the order relations.

It reduces to prove the transitive law. That is to prove ”if \( RF_{x_1} < RF_{x_2} \) and \( RF_{x_2} < RF_{x_3} \), then \( RF_{x_1} < RF_{x_3} \).” We give a simple proof as following.

**Proof:** Let \( x_1, x_2 \) and \( x_3 \) are continuous fuzzy data. We can calculate the central point and radius by Properties 4.1, 4.2 and 4.3. Hence, these expressions result in \( x_1 \equiv (o_1, l_1), x_2 \equiv (o_2, l_2) \) and \( x_3 \equiv (o_3, l_3) \).

Suppose that we have \( RF_{x_1} < RF_{x_2} \) and \( RF_{x_2} < RF_{x_3} \). It means that \( o_1 + [1 - e^{-l_1}] < o_2 + [1 - e^{-l_2}] \) and \( o_2 + [1 - e^{-l_2}] < o_3 + [1 - e^{-l_3}] \). Therefore, \( o_1 - e^{-l_1} < o_2 - e^{-l_2} \) and \( o_2 - e^{-l_2} < o_3 - e^{-l_3} \).
4. IDENTIFYING THE DISTRIBUTION DIFFERENCES OF FUZZY DATA BASED ON A NONPARAMETRIC STATISTICAL METHOD

Hence, we get \( o_1 - e^{-l_1} < o_3 - e^{-l_3} \).
By added one in two sides, we get \( o_1 + [1 - e^{-l_1}] < o_3 + [1 - e^{-l_3}] \).
That is \( RF_{x_1} < RF_{x_3} \). We complete the proof.

Explanation of Definition 4.1

In order to defuzzify the continuous fuzzy data, we define a function \( RF_x \), which is composed of the central point \( o \) and radius \( l \). The concept of the formula (4.1) comes from the mass of a planar lamina (see Definition 3.1). It is a function which can calculate the weight of the area in space. In our case, we use the function to calculate the weight of area between the membership function and \( y \)-axes. In fact that the central point holds the most weight of fuzzy data \( x \). If we do not consider the radius \( l \) in our function \( RF_x \), the function will be defined by only one parametric (central point).

In this chapter, to introduce more information of fuzzy data we added the other parametric (radius). The radius can give a number of weight in the function \( RF_x \). Therefore, we added an increasing function \( 1 - e^{-l} \). This function is combined with an exponential function \( e^{-l} \) which can extend the distance of two data. Hence, we wrote a function \( RF_x \) which is calculated the weight by using central point and radius. Moreover, we rank the fuzzy data by the weight function \( RF_x \).

Now, we define a ranking criterion as follows.

**Definition 4.2 Ranking Criterion**

For \( x_1, x_2 \) are fuzzy data, we define the following ranking criterion.

1. \( RF_{x_1} < RF_{x_2} \) if and only if \( x_1 \prec x_2 \).
2. \( RF_{x_1} = RF_{x_2} \) if and only if \( x_1 \approx x_2 \).
3. \( RF_{x_1} > RF_{x_2} \) if and only if \( x_1 \succ x_2 \).

**Definition 4.3** We say that if \( x_1 \approx x_2 \), it means that \( x_1 \) and \( x_2 \) are in the same classes. Moreover, if \( x_1 \prec x_2 \) or \( x_1 \succ x_2 \), it means that they are in different classes.

For example, let \( x_1 = [5, 8, 16] \) and \( x_2 = [4, 10, 15] \). We can calculate the central point and radius by Property 4.2. Hence we have \( x_1 \equiv (9.67, 2.75) \) and \( x_2 \equiv (9.67, 2.75) \). Moreover, \( RF_{x_1} = RF_{x_2} \). We say that \( x_1 \approx x_2 \) and they are in the same class.
4.2 Identifying the Distribution Difference of Fuzzy Data Based on a Nonparametric Statistical Method

For another example, let $x_1 = [5, 8, 16]$ and $x_2 = [5, 7, 20]$. We can calculate the central point and radius by Property 4.2. Hence we have $x_1 \equiv (9.67, 2.75)$ and $x_2 \equiv (10.67, 3.75)$. Moreover, $RF_{x_1} < RF_{x_2}$. We say that $x_1 \prec x_2$ and they are in different class.

**Definition 4.4** Empirical distribution function with continuous fuzzy value

Let $x_1, x_2, \ldots, x_n$ be $n$ continuous fuzzy data. We can use the function $RF_{x_i}$ to rank the fuzzy data $x_i$ and separate it into different classes $c_i$, which are called Glivenko-Cantelli classes (see discussion in References (13) (15) (61)).

Therefore, we have the order statistic of $x_i$ denoted

$$RF_{x_i(1)} < RF_{x_i(2)} < \ldots < RF_{x_i(n)}$$ (4.2)

Hence, the empirical distribution function can be generalized to a set $C$ to obtain an empirical measure indexed by $c_i$.

$$S_n(c_i) = \frac{1}{n} \sum_{i=1}^{n} I_{c_i}(RF_{x_i}), c_i \in C,$$ (4.3)

where $I_{c_i}$ is the indicator function denoted by

$$I_{c_i}(RF_{x_i}) = \begin{cases} 1, & RF_{x_i} \in c_i, \forall i = 1, 2, \ldots, n. \\ 0, & RF_{x_i} \notin c_i \end{cases}$$ (4.4)

With these definitions, we can now turn to distinguish two populations of fuzzy data based on a nonparametric statistical method by proposing the K-S two-sample test for continuous fuzzy data.

4.2.2 Identifying the Distribution Difference of Fuzzy Data Based on a Nonparametric Statistical Method

In this section, we introduce a nonparametric statistic method for continuous fuzzy data, namely, the K-S two-sample test for continuous fuzzy data. The K-S two-sample test is used to decide whether two independent samples have been drawn from the same population. The test focuses on the agreement between two cumulative distributions. However, how should the K-S two-sample test be adapted for continuous fuzzy data? To address this question, we have developed a new method to derive empirical distribution functions for the continuous fuzzy data in order to find the statistic pivot of the
K-S two-sample test. The procedure of the K-S two-sample test for continuous fuzzy data is as follows:

1. **Samples**: Let \( X_m \) and \( Y_n \) be two samples with continuous fuzzy data. \( X_m \) has size \( m \) and \( Y_n \) has size \( n \). Combining all observations results in \( N = m + n \) data points. A value of the function \( RF \) can be found that will allow us distribute \( X_m \) and \( Y_n \) into different classes \( c_i \), which may be in the same class. The number of classes is less than or equal to \( N \). Moreover, the two empirical distribution functions of \( X_m \) and \( Y_n \) can be derived separately.

2. **Hypothesis**: The two samples have the same distribution \( H_0 \), i.e.

\[
H_0 : F_1(x) = F_2(x)
\]

and

\[
H_1 : F_1(x) \neq F_2(x),
\]

where \( F_1(x) \) denotes the distribution function of one fuzzy sample of size \( m \) and \( F_2(x) \) denotes the distribution function of the other fuzzy sample of size \( n \).

3. **Statistics**: \( D_{m,n} = \sup_X |S_m(X) - S_n(X)| \), where \( S_m(X) \) is the observed cumulative distribution for one sample of size \( m \) and \( S_n(X) \) is the observed cumulative distribution for the other sample of size \( n \).

4. **Decision rule**: The significance level \( \alpha \) is stipulated, and Appendix A Table I is used for making decision.

Similar to the K-S two-sample test with real numbers, the K-S two-sample test with continuous fuzzy data also focuses on the agreement between two cumulative distributions. If the two fuzzy samples have been indeed drawn from the same population, then the cumulative distributions of the fuzzy samples should be close to each other. If the two fuzzy sample cumulative distributions are too far apart at any interval, this implies that the fuzzy samples come from different populations. Therefore, if the two fuzzy sample cumulative distributions have a large deviation, then \( H_0 \) is rejected.
4.3 Empirical Studies

Example 4.1 A Japanese mobile telecommunications company wants to investigate how many times college students send e-mails by mobile phone per day. A manager decides to make a fuzzy questionnaire. A sample was randomly selected of 20 customers (10 males and 10 females) who study at Waseda University in Kitakyushu. The investigator asked them fill the following questions: 1. You send e-mail _____ times (interval) by mobile phone per day. 2. You send e-mail more than _____ times (real numbers) by mobile phone per day. 3. You send e-mail less than _____ times (real numbers) by mobile phone per day. We collected those data and got trapezoidal fuzzy numbers. The answers are shown in Table 4.1.

| Table 4.1: The Numbers of E-mails Per Day by Males and Females |
|----------------|----------------|----------------|----------------|----------------|
| Males          | [0, 6, 9,10]   | [6, 6, 9,10]   | [3, 6, 9,10]   | [5, 6, 8, 8]   |
|                | [5, 5, 8, 8]   | [5, 6, 7, 8]   | [5, 6, 8, 8]   | [5, 7,16,16]   |
|                | [4, 6,12,15]   | [5, 5, 7, 7]   | [5, 5,10,10]   | [15,15,25,30]  |
|                | [5, 7, 8, 8]   | [4, 4, 9,15]   | [5, 5, 7, 8]   | [5, 6,15,20]   |
|                | [5, 7,15,20]   | [5, 7,15,20]   |                |                |
| Females        | [4, 5, 7, 7]   | [5, 5, 7,10]   | [5, 5,10,10]   | [15,15,25,30]  |
|                | [5, 7, 8, 8]   | [4, 4, 9,15]   | [5, 5, 7, 8]   | [5, 6,15,20]   |
|                | [5, 7,15,20]   | [5, 7,15,20]   |                |                |

First, we calculated \( o_i \) and \( l_i \), and then we found the values of \( RF \) and compared them. Moreover, we determined to which class they belong. These calculations are shown in Table 4.2.

After comparing the values of \( RF \), we obtained results in the following inequality:

\[
RF_{Y1} < RF_{Y5} < RF_{X6} < RF_{Y8} < RF_{X1} < RF_{X7} < RF_{X4} = RF_{X8} < RF_{Y6} < RF_{Y2} < \\
RF_{X3} < RF_{X5} < RF_{Y3} < RF_{X2} < RF_{Y7} < RF_{X10} < RF_{X9} < RF_{Y9} < RF_{Y10} < RF_{Y4}
\]

From the above, we derived 19 classes. Then, we found the cumulative distributions of \( X_i \) and \( Y_i \).

From Table 4.3, the following test statistic was obtained:

\[
D = \sup |S_{10}(X) - S_{10}(Y)| = 0.3.
\]

at a significance level of \( \alpha = 0.05 \), \( mnD = 10 \times 10 \times (0.3) = 30 < 70 \) (see Appendix A Table I). Because the observed value did not exceed the critical value, we could not reject \( H_0 \). We conclude that males and females report the same interval for the numbers of times they send e-mail by mobile phone per day.
4. IDENTIFYING THE DISTRIBUTION DIFFERENCES OF FUZZY DATA BASED ON A NONPARAMETRIC STATISTICAL METHOD

Table 4.2: Values for \( o_i, l_i, RF \) and \( C_i \) of RF

<table>
<thead>
<tr>
<th>([a_i, b_i, c_i, d_i])</th>
<th>( o_i )</th>
<th>( l_i )</th>
<th>( RF )</th>
<th>( C_i ) of RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 ) [0, 6, 9,10]</td>
<td>6.03</td>
<td>3.25</td>
<td>6.03 + ( 1 - e^{-3.25} )</td>
<td>5</td>
</tr>
<tr>
<td>( X_2 ) [6, 6, 9,10]</td>
<td>7.76</td>
<td>1.75</td>
<td>7.76 + ( 1 - e^{-1.75} )</td>
<td>13</td>
</tr>
<tr>
<td>( X_3 ) [3, 6, 9,10]</td>
<td>6.93</td>
<td>2.50</td>
<td>6.93 + ( 1 - e^{-2.50} )</td>
<td>10</td>
</tr>
<tr>
<td>( X_4 ) [5, 6, 8, 8]</td>
<td>6.73</td>
<td>1.25</td>
<td>6.73 + ( 1 - e^{-1.25} )</td>
<td>7</td>
</tr>
<tr>
<td>( X_5 ) [5, 5, 8,10]</td>
<td>7.04</td>
<td>2.00</td>
<td>7.04 + ( 1 - e^{-2.00} )</td>
<td>11</td>
</tr>
<tr>
<td>( X_6 ) [5, 5, 8, 8]</td>
<td>6.50</td>
<td>1.50</td>
<td>6.50 + ( 1 - e^{-1.50} )</td>
<td>3</td>
</tr>
<tr>
<td>( X_7 ) [5, 6, 7, 8]</td>
<td>6.50</td>
<td>1.00</td>
<td>6.50 + ( 1 - e^{-1.00} )</td>
<td>6</td>
</tr>
<tr>
<td>( X_8 ) [5, 6, 8, 8]</td>
<td>6.73</td>
<td>1.25</td>
<td>6.73 + ( 1 - e^{-1.25} )</td>
<td>7</td>
</tr>
<tr>
<td>( X_9 ) [5, 7,16,16]</td>
<td>10.98</td>
<td>5.00</td>
<td>10.98 + ( 1 - e^{-5.00} )</td>
<td>16</td>
</tr>
<tr>
<td>( X_{10} ) [4, 6,12,15]</td>
<td>9.27</td>
<td>4.25</td>
<td>9.27 + ( 1 - e^{-4.25} )</td>
<td>15</td>
</tr>
<tr>
<td>( Y_1 ) [4, 5, 7, 7]</td>
<td>5.73</td>
<td>1.25</td>
<td>5.73 + ( 1 - e^{-1.25} )</td>
<td>1</td>
</tr>
<tr>
<td>( Y_2 ) [5, 5, 7,10]</td>
<td>6.86</td>
<td>1.75</td>
<td>6.86 + ( 1 - e^{-1.75} )</td>
<td>9</td>
</tr>
<tr>
<td>( Y_3 ) [5, 5,10,10]</td>
<td>7.50</td>
<td>2.50</td>
<td>7.50 + ( 1 - e^{-2.50} )</td>
<td>12</td>
</tr>
<tr>
<td>( Y_5 ) [5, 5, 7, 7]</td>
<td>6.00</td>
<td>1.00</td>
<td>6.00 + ( 1 - e^{-1.00} )</td>
<td>2</td>
</tr>
<tr>
<td>( Y_6 ) [5, 7, 8, 8]</td>
<td>6.92</td>
<td>1.00</td>
<td>6.92 + ( 1 - e^{-1.00} )</td>
<td>8</td>
</tr>
<tr>
<td>( Y_7 ) [4, 4, 9,15]</td>
<td>8.19</td>
<td>4.00</td>
<td>8.19 + ( 1 - e^{-4.00} )</td>
<td>14</td>
</tr>
<tr>
<td>( Y_8 ) [5, 5, 7, 8]</td>
<td>6.27</td>
<td>1.25</td>
<td>6.27 + ( 1 - e^{-1.25} )</td>
<td>4</td>
</tr>
<tr>
<td>( Y_9 ) [5, 6,15,20]</td>
<td>11.58</td>
<td>6.00</td>
<td>11.58 + ( 1 - e^{-6.00} )</td>
<td>17</td>
</tr>
<tr>
<td>( Y_{10} ) [5, 7,15,20]</td>
<td>11.83</td>
<td>5.75</td>
<td>11.83 + ( 1 - e^{-5.75} )</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4.3: The Cumulative Distributions of \( X_i \) and \( Y_j \)

<table>
<thead>
<tr>
<th>( C_i ) of RF</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{10}(X) )</td>
<td>0</td>
<td>0</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( S_{10}(Y) )</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.3</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>)</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>1</td>
<td>0</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.2</td>
</tr>
</tbody>
</table>
4.3 Empirical Studies

Example 4.2 As the same problem as in Example 4.1, we get the same data in Table 4.1. We used the weight function (WF), which was proposed by Lin, et al. in Reference (40), to defuzzify the fuzzy data. We denoted this method as WF method. The weight function is defined as follows:

$$WF_{x_i} \equiv WF(o_i, l_i) = o_i[1 + ke^{-2l_i}], \forall i = 1, 2, 3, \ldots,$$

where $o_i$ is the central point, $l_i$ is the radius with respect to $o_i$, and $k = max(o_i + l_i) - min(o_j - l_j), \forall i, j = 1, 2, 3,\ldots$

We determined to which class they belong. The results are shown in Table 4.4.

After comparing the values of WF, we obtained results in the following inequality:

$WF_{X_1} < WF_{X_3} < WF_{Y_7} < WF_{Y_3} < WF_{X_{10}} < WF_{X_9} < WF_{Y_9} < WF_{Y_{10}} < WF_{Y_2} < WF_{X_2} < WF_{X_6} < WF_{Y_1} < WF_{Y_8} < WF_{X_4} = WF_{X_8} < WF_{Y_4} < WF_{X_5} < WF_{X_7} < WF_{Y_6}$

From the above, we derived 19 classes. Then, we found the cumulative distributions of $X_i$ and $Y_i$.

From Table 4.5, the following test statistic was obtained:

$$D = sup|S_{10}(X) - S_{10}(Y)| = 0.3.$$  

at a significance level of $\alpha = 0.05$, $mnD = 10 * 10 * (0.3) = 30 < 70$ (see Appendix A Table I). Because the observed value did not exceed the critical value, we could not reject $H_0$. We concluded that males and females report the same interval for the numbers of times they send e-mail by mobile phone per day.

Example 4.3 As the same problem as in Example 4.1, we get the same data in Table 4.1. We used the central point $o_i$ to defuzzify the fuzzy data. This is conventional method. We denoted this method as C method. Moreover, we determined to which class they belong. The results are shown in Table 4.6.

After comparing the values of central point $o_i$, we obtained results in the following inequality:

$C_{Y_1} < C_{Y_5} < C_{X_1} < C_{Y_8} < C_{X_6} = C_{X_7} < C_{X_4} = C_{X_8} < C_{Y_2} < C_{Y_6} < C_{X_3} < C_{X_5} < C_{Y_3} < C_{X_2} < C_{Y_7} < C_{X_{10}} < C_{X_9} < C_{Y_9} < C_{Y_{10}} < C_{Y_4}$

From the above, we derived 18 classes. Then, we found the cumulative distributions of $X_i$ and $Y_i$. 

39
4. IDENTIFYING THE DISTRIBUTION DIFFERENCES OF FUZZY DATA BASED ON A NONPARAMETRIC STATISTICAL METHOD

Table 4.4: Values for $o_i$, $l_i$, $WF$ and $C_i$ of $WF$

<table>
<thead>
<tr>
<th></th>
<th>$[a_i,b_i,c_i,d_i]$</th>
<th>$o_i$</th>
<th>$l_i$</th>
<th>$WF$</th>
<th>$C_i$ of $WF$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>[ 0, 6, 9,10]</td>
<td>6.03</td>
<td>3.25</td>
<td>6.03$[1 + ke^{-6.50}]$</td>
<td>1</td>
</tr>
<tr>
<td>$X_2$</td>
<td>[ 6, 6, 9,10]</td>
<td>7.76</td>
<td>1.75</td>
<td>7.76$[1 + ke^{-3.50}]$</td>
<td>11</td>
</tr>
<tr>
<td>$X_3$</td>
<td>[ 3, 6, 9,10]</td>
<td>6.93</td>
<td>2.50</td>
<td>6.93$[1 + ke^{-5.00}]$</td>
<td>2</td>
</tr>
<tr>
<td>$X_4$</td>
<td>[ 5, 6, 8, 8]</td>
<td>6.73</td>
<td>1.25</td>
<td>6.73$[1 + ke^{-2.50}]$</td>
<td>15</td>
</tr>
<tr>
<td>$X_5$</td>
<td>[ 5, 5, 8,10]</td>
<td>7.04</td>
<td>2.00</td>
<td>7.04$[1 + ke^{-4.00}]$</td>
<td>6</td>
</tr>
<tr>
<td>$X_6$</td>
<td>[ 5, 5, 8, 8]</td>
<td>6.50</td>
<td>1.50</td>
<td>6.50$[1 + ke^{-3.00}]$</td>
<td>12</td>
</tr>
<tr>
<td>$X_7$</td>
<td>[ 5, 6, 7, 8]</td>
<td>6.50</td>
<td>1.00</td>
<td>6.50$[1 + ke^{-2.00}]$</td>
<td>18</td>
</tr>
<tr>
<td>$X_8$</td>
<td>[ 5, 6, 8, 8]</td>
<td>6.73</td>
<td>1.25</td>
<td>6.73$[1 + ke^{-2.50}]$</td>
<td>15</td>
</tr>
<tr>
<td>$X_9$</td>
<td>[ 5, 7,16,16]</td>
<td>10.98</td>
<td>5.00</td>
<td>10.98$[1 + ke^{-10.00}]$</td>
<td>7</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>[ 4, 6,12,15]</td>
<td>9.27</td>
<td>4.25</td>
<td>9.27$[1 + ke^{-8.50}]$</td>
<td>5</td>
</tr>
<tr>
<td>$Y_1$</td>
<td>[ 4, 5, 7, 7]</td>
<td>5.73</td>
<td>1.25</td>
<td>5.73$[1 + ke^{-2.50}]$</td>
<td>13</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>[ 5, 5, 7,10]</td>
<td>6.86</td>
<td>1.75</td>
<td>6.86$[1 + ke^{-3.50}]$</td>
<td>10</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>[ 5, 5,10,10]</td>
<td>7.50</td>
<td>2.50</td>
<td>7.50$[1 + ke^{-5.00}]$</td>
<td>4</td>
</tr>
<tr>
<td>$Y_5$</td>
<td>[ 5, 5, 7, 7]</td>
<td>6.00</td>
<td>1.00</td>
<td>6.00$[1 + ke^{-2.00}]$</td>
<td>17</td>
</tr>
<tr>
<td>$Y_6$</td>
<td>[ 5, 7, 8, 8]</td>
<td>6.92</td>
<td>1.00</td>
<td>6.92$[1 + ke^{-2.00}]$</td>
<td>19</td>
</tr>
<tr>
<td>$Y_7$</td>
<td>[ 4, 4, 9,15]</td>
<td>8.19</td>
<td>4.00</td>
<td>8.19$[1 + ke^{-8.00}]$</td>
<td>3</td>
</tr>
<tr>
<td>$Y_8$</td>
<td>[ 5, 5, 7, 8]</td>
<td>6.27</td>
<td>1.25</td>
<td>6.27$[1 + ke^{-2.50}]$</td>
<td>14</td>
</tr>
<tr>
<td>$Y_9$</td>
<td>[ 5, 6,15,20]</td>
<td>11.58</td>
<td>6.00</td>
<td>11.58$[1 + ke^{-12.00}]$</td>
<td>8</td>
</tr>
<tr>
<td>$Y_{10}$</td>
<td>[ 5, 7,15,20]</td>
<td>11.83</td>
<td>5.75</td>
<td>11.83$[1 + ke^{-11.50}]$</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 4.5: The Cumulative Distributions of $X_i$ and $Y_j$

<table>
<thead>
<tr>
<th>$C_i$ of $WF$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{10}(X)$</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$S_{10}(Y)$</td>
<td>0</td>
<td>0</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>.9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>$</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
</tr>
</tbody>
</table>
### Table 4.6: Values for $C$, and $\bar{C}_i$ of $C$

<table>
<thead>
<tr>
<th></th>
<th>$[a_i,b_i,c_i,d_i]$</th>
<th>$C$</th>
<th>$\bar{C}_i$ of $C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>[0, 6, 9,10]</td>
<td>6.03</td>
<td>6.03</td>
</tr>
<tr>
<td>$X_2$</td>
<td>[6, 6, 9,10]</td>
<td>7.76</td>
<td>7.76</td>
</tr>
<tr>
<td>$X_3$</td>
<td>[3, 6, 9,10]</td>
<td>6.93</td>
<td>6.93</td>
</tr>
<tr>
<td>$X_4$</td>
<td>[5, 6, 8, 8]</td>
<td>6.73</td>
<td>6.73</td>
</tr>
<tr>
<td>$X_5$</td>
<td>[5, 5, 8,10]</td>
<td>7.04</td>
<td>7.04</td>
</tr>
<tr>
<td>$X_6$</td>
<td>[5, 5, 8, 8]</td>
<td>6.50</td>
<td>6.50</td>
</tr>
<tr>
<td>$X_7$</td>
<td>[5, 6, 7, 8]</td>
<td>6.50</td>
<td>6.50</td>
</tr>
<tr>
<td>$X_8$</td>
<td>[5, 6, 8, 8]</td>
<td>6.73</td>
<td>6.73</td>
</tr>
<tr>
<td>$X_9$</td>
<td>[5, 7,16,16]</td>
<td>10.98</td>
<td>10.98</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>[4, 6,12,15]</td>
<td>9.27</td>
<td>9.27</td>
</tr>
<tr>
<td>$Y_1$</td>
<td>[4, 5, 7, 7]</td>
<td>5.73</td>
<td>5.73</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>[5, 5, 7,10]</td>
<td>6.86</td>
<td>6.86</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>[5, 5,10,10]</td>
<td>7.50</td>
<td>7.50</td>
</tr>
<tr>
<td>$Y_5$</td>
<td>[5, 5, 7, 7]</td>
<td>6.00</td>
<td>6.00</td>
</tr>
<tr>
<td>$Y_6$</td>
<td>[5, 7, 8, 8]</td>
<td>6.92</td>
<td>6.92</td>
</tr>
<tr>
<td>$Y_7$</td>
<td>[4, 4, 9,15]</td>
<td>8.19</td>
<td>8.19</td>
</tr>
<tr>
<td>$Y_8$</td>
<td>[5, 5, 7, 8]</td>
<td>6.27</td>
<td>6.27</td>
</tr>
<tr>
<td>$Y_9$</td>
<td>[5, 6,15,20]</td>
<td>11.58</td>
<td>11.58</td>
</tr>
<tr>
<td>$Y_{10}$</td>
<td>[5, 7,15,20]</td>
<td>11.83</td>
<td>11.83</td>
</tr>
</tbody>
</table>

### Table 4.7: The Cumulative Distributions of $X_i$ and $Y_j$

| $S_{10}(X)$ | $S_{10}(Y)$ | $|S_{10}(X) - S_{10}(Y)|$ |
|-------------|-------------|-------------------------|
| 0           | 0.1         | 0.1                     |
| 0.1         | 0.2         | 0.1                     |
| 0.2         | 0.3         | 0.1                     |
| 0.3         | 0.4         | 0.1                     |
| 0.4         | 0.5         | 0.1                     |
| 0.5         | 0.6         | 0.1                     |
| 0.6         | 0.7         | 0.1                     |
| 0.7         | 0.8         | 0.1                     |
| 0.8         | 0.9         | 0.1                     |
| 0.9         | 1.0         | 0.1                     |

41
4. IDENTIFYING THE DISTRIBUTION DIFFERENCES OF FUZZY DATA BASED ON A NONPARAMETRIC STATISTICAL METHOD

From Table 4.7, the following test statistic was obtained:

\[ D = \sup |S_{10}(X) - S_{10}(Y)| = 0.3. \]

at a significance level of \( \alpha = 0.05 \), \( mnD = 10 \times 10 \times (0.3) = 30 < 70 \) (see Appendix A Table I). Because the observed value did not exceed the critical value, we could not reject \( H_0 \). We concluded that males and females report the same interval for the numbers of times they send e-mail by mobile phone per day.

Example 4.4 As the same problem as in Example 4.1, we get the same data in Table 4.1. We use the distance method developed by Cheng (8). We denoted this method as \( RD \) method. We give the result in Table 4.8.

Note that we have the membership function \( f(x) \) as in Property 4.3. Here \( \bar{x}_i = o_i \) and \( \bar{y}_i = \int_0^1 g_L^Y dy + \int_0^1 g_R^Y dy \int_1^0 g_L^X dy + \int_1^0 g_R^X dy \int_1^0 g_L^X dy + \int_1^0 g_R^X dy = (a + d) + 2(b + c) \), where \( g_L^X = a + (b - a)y \) and \( g_R^X = d + (c - d)y \) are inverse functions \( \forall y \in [0, 1] \). Moreover, \( R(\tilde{X}_i) = \sqrt{\bar{x}_i^2 + \bar{y}_i^2} \).

After comparing the values of \( R(\tilde{X}_i) \), we obtained results in the following inequality:

\[ RD_{Y_1} < RD_{Y_5} < RD_{X_1} < RD_{Y_8} < RD_{X_6} = RD_{X_7} < RD_{X_4} = RD_{X_8} < RD_{Y_2} < RD_{Y_6} < RD_{X_3} < RD_{X_5} < RD_{Y_3} < RD_{X_2} < RD_{Y_7} < RD_{X_{10}} < RD_{X_9} < RD_{Y_9} < RD_{Y_{10}} < RD_{Y_4} \]

From the above, we derived 18 classes. Then, we found the cumulative distributions of \( X_i \) and \( Y_i \).

From Table 4.9, the following test statistic was obtained:

\[ D = \sup |S_{10}(X) - S_{10}(Y)| = 0.3. \]

at a significance level of \( \alpha = 0.05 \), \( mnD = 10 \times 10 \times (0.3) = 30 < 70 \) (see Appendix A Table I). Because the observed value did not exceed the critical value, we could not reject \( H_0 \). We conclude that males and females report the same interval for the numbers of times they send e-mail by mobile phone per day.

4.4 Comparison of Each Methods

To clarify the results of each method with K-S two sample test, we gave the values of four defuzzification formulas and classifications \( C_i \) in Table 4.10. Moreover, we gave
4.4 Comparison of Each Methods

Table 4.8: Values for $\bar{x}_i$, $\bar{y}_i$, $RD$ and $C_i$ of $RD$

<table>
<thead>
<tr>
<th>$[a_i,b_i,c_i,d_i]$</th>
<th>$\bar{x}_i$</th>
<th>$\bar{y}_i$</th>
<th>$RD$</th>
<th>$C_i$ of $RD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ [0, 6, 9,10]</td>
<td>6.03</td>
<td>0.53</td>
<td>$\sqrt{(6.03)^2 + (0.53)^2}$</td>
<td>3</td>
</tr>
<tr>
<td>$X_2$ [6, 6, 9,10]</td>
<td>7.76</td>
<td>0.49</td>
<td>$\sqrt{(7.76)^2 + (0.49)^2}$</td>
<td>12</td>
</tr>
<tr>
<td>$X_3$ [3, 6, 9,10]</td>
<td>6.93</td>
<td>0.51</td>
<td>$\sqrt{(6.93)^2 + (0.51)^2}$</td>
<td>9</td>
</tr>
<tr>
<td>$X_4$ [5, 6, 8, 8]</td>
<td>6.73</td>
<td>0.51</td>
<td>$\sqrt{(6.73)^2 + (0.51)^2}$</td>
<td>6</td>
</tr>
<tr>
<td>$X_5$ [5, 5, 8,10]</td>
<td>7.04</td>
<td>0.49</td>
<td>$\sqrt{(7.04)^2 + (0.49)^2}$</td>
<td>10</td>
</tr>
<tr>
<td>$X_6$ [5, 5, 8, 8]</td>
<td>6.50</td>
<td>0.50</td>
<td>$\sqrt{(6.50)^2 + (0.50)^2}$</td>
<td>5</td>
</tr>
<tr>
<td>$X_7$ [5, 6, 7, 8]</td>
<td>6.50</td>
<td>0.50</td>
<td>$\sqrt{(6.50)^2 + (0.50)^2}$</td>
<td>5</td>
</tr>
<tr>
<td>$X_8$ [5, 6, 8, 8]</td>
<td>6.73</td>
<td>0.51</td>
<td>$\sqrt{(6.73)^2 + (0.51)^2}$</td>
<td>6</td>
</tr>
<tr>
<td>$X_9$ [5, 7,16,16]</td>
<td>10.98</td>
<td>0.51</td>
<td>$\sqrt{(10.98)^2 + (0.51)^2}$</td>
<td>15</td>
</tr>
<tr>
<td>$X_{10}$ [4, 6,12,15]</td>
<td>9.27</td>
<td>0.50</td>
<td>$\sqrt{(9.27)^2 + (0.50)^2}$</td>
<td>14</td>
</tr>
<tr>
<td>$Y_1$ [4, 5, 7, 7]</td>
<td>5.73</td>
<td>0.51</td>
<td>$\sqrt{(5.73)^2 + (0.51)^2}$</td>
<td>1</td>
</tr>
<tr>
<td>$Y_2$ [5, 5, 7,10]</td>
<td>6.86</td>
<td>0.48</td>
<td>$\sqrt{(6.86)^2 + (0.48)^2}$</td>
<td>7</td>
</tr>
<tr>
<td>$Y_3$ [5, 5,10,10]</td>
<td>7.50</td>
<td>0.50</td>
<td>$\sqrt{(7.50)^2 + (0.50)^2}$</td>
<td>11</td>
</tr>
<tr>
<td>$Y_4$ [15,15,25,30]</td>
<td>21.33</td>
<td>0.49</td>
<td>$\sqrt{(21.33)^2 + (0.49)^2}$</td>
<td>18</td>
</tr>
<tr>
<td>$Y_5$ [5, 5, 7, 7]</td>
<td>6.00</td>
<td>0.50</td>
<td>$\sqrt{(6.00)^2 + (0.50)^2}$</td>
<td>2</td>
</tr>
<tr>
<td>$Y_6$ [5, 7, 8, 8]</td>
<td>6.92</td>
<td>0.51</td>
<td>$\sqrt{(6.92)^2 + (0.51)^2}$</td>
<td>8</td>
</tr>
<tr>
<td>$Y_7$ [4, 4, 9,15]</td>
<td>8.19</td>
<td>0.48</td>
<td>$\sqrt{(8.19)^2 + (0.48)^2}$</td>
<td>13</td>
</tr>
<tr>
<td>$Y_8$ [5, 5, 7, 8]</td>
<td>6.27</td>
<td>0.49</td>
<td>$\sqrt{(6.27)^2 + (0.49)^2}$</td>
<td>4</td>
</tr>
<tr>
<td>$Y_9$ [5, 6,15,20]</td>
<td>11.58</td>
<td>0.49</td>
<td>$\sqrt{(11.58)^2 + (0.49)^2}$</td>
<td>16</td>
</tr>
<tr>
<td>$Y_{10}$ [5, 7,15,20]</td>
<td>11.83</td>
<td>0.49</td>
<td>$\sqrt{(11.83)^2 + (0.49)^2}$</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 4.9: The Cumulative Distributions of $X_i$ and $Y_j$

<table>
<thead>
<tr>
<th>$C_i$ of $RD$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{10}(X)$</td>
<td>0</td>
<td>0</td>
<td>.1</td>
<td>.1</td>
<td>.3</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$S_{10}(Y)$</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.3</td>
<td>.3</td>
<td>.4</td>
<td>.5</td>
<td>.5</td>
<td>.6</td>
<td>.6</td>
<td>.7</td>
<td>.7</td>
<td>.7</td>
<td>.8</td>
<td>.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$</td>
<td>S_{10}(X) - S_{10}(Y)</td>
<td>$</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>0</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
<td>.1</td>
<td>.2</td>
<td>.1</td>
<td>.2</td>
<td>.3</td>
<td>.2</td>
<td>.1</td>
<td>0</td>
</tr>
</tbody>
</table>
4. IDENTIFYING THE DISTRIBUTION DIFFERENCES OF FUZZY DATA BASED ON A NONPARAMETRIC STATISTICAL METHOD

the comparison results of each method in Table 4.11. From Table 4.11, we could see that we got the same statistic and decision of K-S two sample test in all methods. The differences between the methods are that the defuzzification formulas and the number of classes are different. The RF and WF methods are more than one class than C and RD methods. It means that RF and RD methods are more effective to classify the fuzzy data because they could divide the fuzzy data into more subdivisions.

To see the value $c_i$ of C and $c_i$ of RD in Table 4.10, C method and RD method obtained the same value. It means that although Cheng (RD method) used two variables, $\tilde{X}_i$ (the value on the horizontal axis) and $\tilde{Y}_i$ (the value on the vertical axis), to determine the RD value, but the value $\tilde{Y}_i$ did not give any effectiveness in calculating the value of $c_i$. Hence, we thought that this method (RD method) did not have any contribution than conventional method (C method).

From Table 4.10, we could also see that the value of $c_i$ of WF was decreasing while the other three values of $c_i$ ($c_i$ of RF, $c_i$ of C and $c_i$ of RD) are all increasing.

For example, we took three trapezoidal fuzzy numbers, $X_5$, $X_7$ and $X_8$, the order among each method is given as follows.

$$RF_{X_7} < RF_{X_8} < RF_{X_5} : 6 < 7 < 11$$

$$WF_{X_7} > WF_{X_8} > WF_{X_5} : 18 > 15 > 6$$

$$C_{X_7} < C_{X_8} < C_{X_5} : 5 < 6 < 10$$

$$RD_{X_7} < RD_{X_8} < RD_{X_5} : 5 < 6 < 10$$

It means that when we use two variables (central point $o_i$ and radius $r_i$) in WF method, the value of weight function is decreasing when the value of $r_i$ is increasing. It is not satisfied the Raking Criterion in Definition 4.2. Moreover, we can not calculate the order statistic in Definition 4.4 by using this weight function.

Although WF and RF methods are all in the same results and also can divide the fuzzy data into more subdivisions, we thought that it is not suitable to use WF method with K-S two sample test here because the defuzzification formula is not satisfied by Definitions 4.2 and 4.4. Hence, we conclude that our method (RF method) is more effective and successful with K-S two sample test.
4.4 Comparison of Each Methods

Table 4.10: Values for RF, εi of RF, WF, εi of WF, C, εi of C, RD and εi of RD

<table>
<thead>
<tr>
<th>X</th>
<th>RF</th>
<th>εi of RF</th>
<th>WF</th>
<th>εi of WF</th>
<th>C</th>
<th>εi of C</th>
<th>RD</th>
<th>εi of RD</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>6.9912</td>
<td>5</td>
<td>6.2548</td>
<td>1</td>
<td>6.03</td>
<td>3</td>
<td>6.0532</td>
<td>3</td>
</tr>
<tr>
<td>X2</td>
<td>8.5862</td>
<td>13</td>
<td>13.5714</td>
<td>11</td>
<td>7.76</td>
<td>12</td>
<td>7.7755</td>
<td>12</td>
</tr>
<tr>
<td>X3</td>
<td>7.8479</td>
<td>10</td>
<td>8.0880</td>
<td>2</td>
<td>6.93</td>
<td>9</td>
<td>6.9487</td>
<td>9</td>
</tr>
<tr>
<td>X4</td>
<td>7.4435</td>
<td>7</td>
<td>20.4303</td>
<td>15</td>
<td>6.73</td>
<td>6</td>
<td>6.7493</td>
<td>6</td>
</tr>
<tr>
<td>X5</td>
<td>7.9047</td>
<td>11</td>
<td>10.2378</td>
<td>6</td>
<td>7.04</td>
<td>10</td>
<td>7.0570</td>
<td>10</td>
</tr>
<tr>
<td>X6</td>
<td>7.2769</td>
<td>3</td>
<td>14.5257</td>
<td>12</td>
<td>6.50</td>
<td>5</td>
<td>6.5192</td>
<td>5</td>
</tr>
<tr>
<td>X7</td>
<td>7.1321</td>
<td>6</td>
<td>28.3161</td>
<td>18</td>
<td>6.50</td>
<td>5</td>
<td>6.5192</td>
<td>5</td>
</tr>
<tr>
<td>X8</td>
<td>7.4435</td>
<td>7</td>
<td>20.4303</td>
<td>15</td>
<td>6.73</td>
<td>6</td>
<td>6.7493</td>
<td>6</td>
</tr>
<tr>
<td>X9</td>
<td>11.9733</td>
<td>16</td>
<td>10.9924</td>
<td>7</td>
<td>10.98</td>
<td>15</td>
<td>10.9918</td>
<td>15</td>
</tr>
<tr>
<td>X10</td>
<td>10.2557</td>
<td>15</td>
<td>9.3168</td>
<td>5</td>
<td>9.27</td>
<td>14</td>
<td>9.2835</td>
<td>14</td>
</tr>
<tr>
<td>Y1</td>
<td>6.4435</td>
<td>1</td>
<td>17.3946</td>
<td>13</td>
<td>5.73</td>
<td>1</td>
<td>5.7527</td>
<td>1</td>
</tr>
<tr>
<td>Y2</td>
<td>7.6862</td>
<td>9</td>
<td>11.9974</td>
<td>10</td>
<td>6.86</td>
<td>7</td>
<td>6.8768</td>
<td>7</td>
</tr>
<tr>
<td>Y3</td>
<td>8.4179</td>
<td>12</td>
<td>8.7533</td>
<td>4</td>
<td>7.50</td>
<td>11</td>
<td>7.5166</td>
<td>11</td>
</tr>
<tr>
<td>Y5</td>
<td>6.6321</td>
<td>2</td>
<td>26.1379</td>
<td>17</td>
<td>6.00</td>
<td>2</td>
<td>6.0208</td>
<td>2</td>
</tr>
<tr>
<td>Y6</td>
<td>7.5521</td>
<td>8</td>
<td>30.1457</td>
<td>19</td>
<td>6.92</td>
<td>8</td>
<td>6.9388</td>
<td>8</td>
</tr>
<tr>
<td>Y7</td>
<td>9.1717</td>
<td>14</td>
<td>8.2581</td>
<td>3</td>
<td>8.19</td>
<td>13</td>
<td>8.2041</td>
<td>13</td>
</tr>
<tr>
<td>Y8</td>
<td>6.9835</td>
<td>4</td>
<td>19.0339</td>
<td>14</td>
<td>6.27</td>
<td>4</td>
<td>6.2891</td>
<td>4</td>
</tr>
<tr>
<td>Y9</td>
<td>12.5775</td>
<td>17</td>
<td>11.5818</td>
<td>8</td>
<td>11.58</td>
<td>16</td>
<td>11.5904</td>
<td>16</td>
</tr>
<tr>
<td>Y10</td>
<td>12.8268</td>
<td>18</td>
<td>11.8330</td>
<td>9</td>
<td>11.83</td>
<td>17</td>
<td>11.8401</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 4.11: The Results of K-S Two Sample Test in Each Method

<table>
<thead>
<tr>
<th>Defuzzification</th>
<th>Formula</th>
<th>Statistic</th>
<th>Classes</th>
<th>D10,10</th>
<th>10 * 10 * D10,10 V.S P0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF Method</td>
<td>o + [1 - e^{-1}]</td>
<td></td>
<td>19</td>
<td>0.3</td>
<td>30 &lt; 70</td>
</tr>
<tr>
<td>WF Method</td>
<td>o × [1 + ke^{-2l}]</td>
<td></td>
<td>19</td>
<td>0.3</td>
<td>30 &lt; 70</td>
</tr>
<tr>
<td>C Method</td>
<td>o</td>
<td></td>
<td>18</td>
<td>0.3</td>
<td>30 &lt; 70</td>
</tr>
<tr>
<td>RD Method</td>
<td>R(\tilde{x}_i)</td>
<td></td>
<td>18</td>
<td>0.3</td>
<td>30 &lt; 70</td>
</tr>
</tbody>
</table>

* P0.05 is a critical value under significant level α = 0.05.
4. IDENTIFYING THE DISTRIBUTION DIFFERENCES OF FUZZY DATA BASED ON A NONPARAMETRIC STATISTICAL METHOD

4.5 Summary of Chapter 4

In various research studies, many evaluations and psychological tests are conducted and many decisions are made using surveys and/or questionnaires that seek people’s opinions. It is routine to ask people about their opinions according to binary, multiple-choice questions, but people actually have complex and/or vague thoughts. If we want to understand human’s thoughts in reality, we must extend classical statistical models to address fuzzy numbers.

To identify the distribution differences between two populations of fuzzy data, we presented a new function called $RF_x$ in this chapter, which consists of both central points and radius values. The function $RF_x$ can classify all of the continuous fuzzy data in a dataset. Moreover, it allows us to differentiate two fuzzy samples. As a result, a cumulative distribution function can be determined and we can obtain the statistical pivot of the K-S two-sample test with continuous fuzzy data. We also described several empirical studies to compare the proposed method with conventional methods. In the $WF$, $RD$ and $RF$ methods, fuzzy data are ranked based not only on one point but also on another parameter. The difference between these three methods is that Cheng ($RD$ method) chose the central point by calculating the moment about the $x$-value and the other parameter ($y$-value) based on an inverse function, but we ($WF$ and $RF$ methods) chose the central point by calculating the moment about the $x$-value and the other parameter (radius) by calculating the area between membership function and the $x$-value.

In this chapter, we have provided a method of classifying the continuous fuzzy data that enables us to use the K-S two-sample test to identify the distribution differences between two populations. Through this procedure, an intelligent calculation method can be applied to analyzing industrial, physiological, economic or financial data in the future.
Risk Assessment of a Portfolio Selection Model Based on a Fuzzy Statistical Test

5.1 Introduction

The portfolio selection model has been well developed on the basis of a mean-variance approach. It was first proposed by Markowitz (44),(45),(46), who combined probability and optimization theories to analyze the performance of economic agents. The key principle of the mean-variance model is to use the expected return of a portfolio as the investment return and the variance of the expected returns of the portfolio as the investment risk. Most of the existing portfolio selection models are based on probability theory. The mean-variance portfolio selection problem has been studied by Sharpe (62), Merton (48), Perold (56), Pang (54), Voros (70), and Best (2), (3). To analyze uncertain phenomena in the real world, multivariate data analysis has been applied to investigate portfolio selection problems. For example, Zhang et al. (89) discussed the portfolio selection problem when the returns of assets are fuzzy numbers. Hasuike et al. (18) discussed two portfolio selection problems including probabilistic future and ambiguous expected returns. Moreover, Hasuike and Ishii (17) discussed a portfolio selection problem with type-2 fuzzy returns involving interval numbers based on the investor’s subjectivity. Giove et al. (16) discussed a portfolio selection problem in which the prices of the securities are treated as interval variables. To deal with such an
interval portfolio problem, they adopted a minimax regret approach based on a regret function.

Some other research works have discussed how to solve fuzzy portfolio selection models, such as Peng et al. (55), who addressed the portfolio selection problems in fuzzy environments by employing a credibility programming approach based on a credibility measure (see (42)). Wang et al. (71) proposed a new real options analysis approach by combining the binomial lattice-based model with a fuzzy random variable. Zhang et al. (86) proposed a model to convert the forecasted uncertain values into normal fuzzy numbers. Tanaka et al. (see the discussion in (65) and (66)) proposed two types of portfolio selection models based on fuzzy probabilities and exponential possibility distributions, respectively. Wang and Zhu (74), and Lai et al. (32) constructed interval programming models of portfolio selection. Zhang and Wang (88) and Zhang et al. (87) discussed the portfolio selection problem based on the (crisp) possibilistic mean and variance when short sales are not allowed for all risky assets. Watada (75), Ramaswamy (59), and Leon et al. (36) discussed portfolio selection using fuzzy decision theory.

Most studies have not considered any type of probability distribution function with fuzzy random variables. Moreover, no statistical test has been applied to examine the results of the portfolio selection model with fuzzy data. In view of these weaknesses, the objective of this paper is to develop a statistical test to evaluate the results of the portfolio selection model with fuzzy data. First, we address the problem of finding the distribution function with fuzzy data.

The distribution function must be predicted under a specified condition or for a situation known in advance (see (50)). When we want to work with fuzzy data, the underlying probability distribution of the fuzzy data is not known. It is not easy to describe such information in terms of statistics. Therefore, we must establish techniques to handle such information and knowledge. Following Zadeh (83) (84), we use fuzzy set theory and take the concept of fuzzy statistics into consideration. Fundamental statistical measurements such as the mean, the median and the mode are useful for illustrating the characteristics of a sample distribution. More research works should focus on the fuzzy statistical aspects of the model and their applications in engineering, medical and social science. Wu and Cheng (78) identified a model structure through qualitative simulation; Casalino et al. (5), Esogbue and Song (11), and Wu and Sun (79) discussed the concepts of fuzzy statistics and applied them to social surveys. Wu
and Tseng (80) used the fuzzy regression method of coefficient estimation to analyze the Taiwan monitoring index of economics. All of the above-mentioned studies dealt with problems by using central point values. Lin et al. (40) proposed a new weight function for fuzzy numbers defined by the central point and the radius. Moreover, Lin et al. (37) proposed a method to recognize the underlying distribution function using its central point and radius, which gives us more information about the original fuzzy data.

The objective of this chapter is to build a statistical test of fuzzy data and apply it to a portfolio selection problem with interval values, and then to statistically evaluate the best return. In the first step, we need to identify the probability distribution function and each parameter in the probability distribution function. When we know the distribution function of each parameter, we can easily calculate the expected return and variance. Those values can enable us to define a portfolio selection model with interval values. We can also make a decision based on a fuzzy statistical test that statistically determines and explicitly states whether we should accept the risk in the investment.

The rest of this chapter consists of the following: Section 5.2 provides a brief review of related studies. The main method is described in Section 5.3. Section 5.4 illustrates empirical studies with interval values and how to apply a fuzzy statistical test to the portfolio selection model. Finally, concluding remarks and suggested topics for further studies are presented in Section 5.5.

5.2 Notations and Preliminary Definitions

To proceed with the detailed discussion, it would be more advantageous to introduce some useful notation.

Note that fuzzy number $F_i$ denotes a vector form because of central point and radius.

**Notation**

- $A_i$: acceptance region of $i$
- $\mathcal{A}$: 2-dimensional acceptance region $\mathcal{A} = A_o \times A_l$
- $\alpha$: significance level
- $C$: a subset of a specified collection of elements in $U$
5. RISK ASSESSMENT OF A PORTFOLIO SELECTION MODEL BASED ON A FUZZY STATISTICAL TEST

$C_i$: a specified collection of element $i$  
$\varepsilon$: risk level  
$F_i$: interval values $F_i = [a_i, b_i]$  
$\bar{F}$: fuzzy sample mean  
$I$: unit vector $I = [1, 1, \cdots, 1]'$  
$I_{C_i}$: characteristic function of $C_i$  
k: risk level  
$K$: decision set of $k$  
l_i: the $i$ radius values of $F_i$  
$L^*$: optimal solution in model (5.15)  
$L$: a stochastic quantity with mean $m_l$ and variance $\sigma_l^2$  
m_0: a specified value  
m_i: mean values of $i$  
m_{i,o}: a specified value of $i_0$  
m: mean vector $m \equiv (m_o, m_l)$  
m_0: a specified vector $m_0 \equiv (m_{o,0}, m_{l,0})$  
$\mu$: membership function  
o_i: the $i$ central point values of $F_i$  
$O$: a stochastic quantity with mean $m_o$ and variance $\sigma_o^2$  
$\Omega$: probability space  
r_i: the return rate of asset $i$  
r: the return vector $r = [r_1, r_2, \cdots, r_n]'$  
$\bar{r}$: the expected return vector $\bar{r} = [\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_n]'$  
$R$: return of portfolio vector $x$  
$\Re$: real numbers  
$\Sigma$: a $2 \times 2$ covariance matrix $\Sigma = [\text{cov}(O, L)]_{2\times2}$  
$S_{n,i}$: sample standard deviation of the data set with size $n$  
$S_n$: standard deviation vector $S_n \equiv (S_{n,o}, S_{n,l})$  
$\sigma_{ij}^2$: variance of row $i$ and column $j$  
t: variables in $t$-distribution  
$T_i$: the $i$ statistic for the hypothesis $H_0$  
$T$: a statistic vector $T \equiv (T_o, T_l)$ for the hypothesis $H_0$  
$U$: universal set
5.2 Notations and Preliminary Definitions

$U^*$: optimal solution in model (5.16)

$V$: an $n \times n$ covariance matrix $V = [\sigma_{ij}]_{n \times n}$

$W_n$: sum of $X_i$ with size $n$

$x$: a portfolio vector $x = [x_1, x_2, \cdots, x_n]'$

$x^*$: optimal solution vector of model (5.1) which is $x^* = [x^*_1, x^*_2, \cdots, x^*_n]'$

$x_i$: the proportion invested in asset $i$

$X_i$: random variables of $i$

$\bar{X}_n$: sample mean of the data set with size $n$

We have defined a method to calculate the central point $o$, and radius $l_i$ of interval value in Property 3.1. We rewrite the property in the following definition.

**Definition 5.1** An interval value is denoted as $F = [a, b]$ with a central point $o = \frac{a + b}{2}$ and radius $l = \frac{b - a}{2}$. We give the notation as $F \equiv (o, l)$.

Let us recall that a fuzzy statistical test means a statistical test which can deal with fuzzy data. First, we introduce a traditional statistical test in the following subsection.

5.2.1 Statistical Analysis

Let $X_1, \cdots, X_n$ be a sequence of random variables (not necessarily normally distributed). We say that the $X_i$ are independently identically distributed (i.i.d) if the $X_i$ are independent and have the same distribution. We write $W_n = \sum_{i=1}^{n} X_i$ and $\bar{X}_n = \frac{W_n}{n}$

to denote the total and average, respectively, of the $nX_i's$. We introduce the most important theorem in statistics as follows.

**Theorem 5.1 Central Limit Theorem (1)**

Let $X_1, \cdots, X_n$ be i.i.d. random variables with mean $m$ and variance $\sigma^2$. Let

$$Z_n = \frac{\frac{1}{n}(\bar{X}_n - m)}{\sigma} = \frac{W_n - nm}{n^{\frac{1}{2}} \sigma}.$$ 

Then, $Z_n$ converges in distribution to $Z$ as $n \to \infty$. We denote

$$Z_n \to Z \sim N(0, 1) \quad as \quad n \to \infty.$$
where \( Z \) distributes according to a standard normal distribution function \( N(0, 1) \).

Note that \( m \) denotes mean and \( \mu \) expresses a membership function.

In statistics, functions of observations are often important. Therefore, we introduce the following statistical test.

**Theorem 5.2**  
**T-test** (69)

For samples \( X_1, X_2, \cdots, X_n \) of a normally distributed stochastic quantity \( X \sim N(m, \sigma^2) \), a statistic for the hypothesis \( H_0 : m = m_0 \) is

\[
T = \frac{\bar{X}_n - m_0}{S_n \sqrt{n}},
\]

and the acceptance region \( A \) for \( T \) under probability \( \alpha \) for an error of the first type is

\[
A = \{ t \in \mathbb{R} | t = | \frac{\bar{X}_n - m_0}{S_n \sqrt{n}} | \leq t_{n-1,1-\frac{\alpha}{2}} \},
\]

where \( \bar{X}_n \) and \( S_n \) are the sample mean and the sample standard deviation of the data, \( m_0 \) is a specified value, \( n \) the sample size and \( t_{n-1,1-\frac{\alpha}{2}} \) the \((1 - \frac{\alpha}{2})\)-fractile of the \( t \)-distribution with degree of freedom \( n - 1 \), respectively.

When the above information is given, we have provided all prior knowledge for dealing with a portfolio selection problem.

Now, we give a brief description of the portfolio selection model in the following subsection.

### 5.2.2 Markowitz’s Portfolio Selection Model

Markowitz’s mean-variance model is based on a probability distribution in which uncertainty is equated with randomness (44) (45) (46). That is, the return on the \( i \)th asset, \( r_i \), will be regarded as a random variable.

Consider a market with \( n \) risky assets. An investor’s position in this market is described by a portfolio vector \( \mathbf{x} = [x_1, x_2, \cdots, x_n]' \), where the \( i \)th component \( x_i \) represents the proportion invested in asset \( i \). The return vector on portfolio vector \( \mathbf{x} \) is denoted \( \mathbf{r} = [r_1, r_2, \cdots, r_n]' \), where \( r_i \) represents the return rate of asset \( i \). In the conventional mean-variance methodology for portfolio selection, \( r_i \) is regarded as a random
variable, \( \forall i = 1, 2, \ldots, n \). Let \( \bar{r} = [\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_n]' \) and \( \mathbf{V} = [\sigma_{ij}]_{n \times n} \) be the expected return vector and covariance matrix, respectively. The return \( R \) on the portfolio \( \mathbf{x} \) is given by \( R = \sum_{i=1}^{n} r_i x_i \). Set \( \mathbf{I} = [1, 1, \cdots, 1]' \). The objective of the investor is to choose a portfolio that maximizes the return on the investment under some constraints on the risk of the investment.

A mean-variance model can be formulated mathematically for portfolio selection as

\[
\begin{align*}
\text{max} & \quad \bar{r} \mathbf{x} \\
\text{s.t.} & \quad \sqrt{\mathbf{x}' \mathbf{V} \mathbf{x}} \leq \varepsilon \\
& \quad \sum_{i=1}^{n} x_i \leq 1 \\
& \quad x_i \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{align*}
\]

(5.3)

where \( \varepsilon (\varepsilon \geq 0) \) represents the risk level, \( \mathbf{x}' \mathbf{V} \mathbf{x} = \sum_{i=1}^{n} \sigma_{ii}^2 x_i^2 + \sum_{j=1}^{n} \sum_{i=1, i\neq j}^{n} \sigma_{ij} x_i x_j \), \( \sigma_{ij} = \text{cov}(a_i, a_j) \) is the covariance, and \( a_i \) and \( a_j \) are random variables, \( \forall i, j = 1, 2, \cdots, n \).

Here, we rewrite formula (5.3) clearly in the following formula:

\[
\begin{align*}
\text{max} & \quad E\left(\sum_{i=1}^{n} r_i x_i\right) \\
\text{s.t.} & \quad \sqrt{\mathbf{x}' \mathbf{V} \mathbf{x}} \leq \varepsilon \\
& \quad \sum_{i=1}^{n} x_i \leq 1 \\
& \quad x_i \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{align*}
\]

(5.4)

That is, we need to solve the following programming problem:

\[
\begin{align*}
\text{max} & \quad \sum_{i=1}^{n} E(r_i) x_i \\
\text{s.t.} & \quad \sqrt{\mathbf{x}' \mathbf{V} \mathbf{x}} \leq \varepsilon \\
& \quad \sum_{i=1}^{n} x_i \leq 1 \\
& \quad x_i \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{align*}
\]

(5.5)

We will introduce the fuzzy statistical test on portfolio selection model in the following section.

### 5.3 Fuzzy Statistical Test on the Portfolio Selection Model

Assume that there are \( n \) distinct tradable assets in the market. The terminal rate of return for asset \( i \), denoted as \( r_i, \forall i = 1, 2, \cdots, n \), is assumed to be a fuzzy random
variable. By the widely accepted definition, the expectation of a fuzzy random variable is a fuzzy variable. We give some definitions to express the total fuzzy return $R$ on a portfolio vector $x$, where $x = [x_1, x_2, \cdots, x_n]'$ is an $n$ row vector.

5.3.1 Portfolio Selection Model with Interval Values

Let $F_i = [a_i, b_i] \equiv (o_i, l_i)$, $\forall i = 1, 2, \cdots, n$, be interval values on the probability space $\Omega$, where $o_i$ is a random variable of central point of $F_i$, and $l_i$ is a random variable of the radius of $F_i$, $\forall i = 1, 2, \cdots, n$.

**Definition 5.2 Fuzzy Expected Return (data with interval values)**

For the proportion invested in asset $i$, $\forall i = 1, 2, \cdots, n$, we have $x = [x_1, x_2, \cdots, x_n]'$. Definition 5.1 specifies the fuzzy expected return of $F_i$, i.e. $F_i \equiv (o_i, l_i)$, $\forall i = 1, 2, \cdots, n$, as follows:

$$E[R(x)] = \sum_{i=1}^{n} E(F_i) x_i = \left( \sum_{i=1}^{n} E(o_i) x_i \right) \left( \sum_{i=1}^{n} E(l_i) x_i \right).$$

Note that $R[x] = \sum_{i=1}^{n} F_i x_i$, $\forall i = 1, 2, \cdots, n$.

**Definition 5.3 Fuzzy Portfolio Variance (data with interval values)**

For the proportion invested in asset $i$, $\forall i = 1, 2, \cdots, n$. We have $x = [x_1, x_2, \cdots, x_n]'$. Definition 5.1 specifies the fuzzy portfolio variance of $F_i$, $\forall i = 1, 2, \cdots, n$, as follows:

$$\text{var}[R(x)] = \sum_{i=1}^{n} \text{var}(F_i) x_i = \left( \text{var} \left( \sum_{i=1}^{n} o_i x_i \right) \right) \left( \text{var} \left( \sum_{i=1}^{n} l_i x_i \right) \right),$$

where

$$\text{var} \left( \sum_{i=1}^{n} o_i x_i \right) = \sum_{i=1}^{n} \sigma^2_{o_i} x_i^2 + \sum_{j=1}^{n} \sum_{i=1, i \neq j}^{n} \sigma_{o_i o_j} x_i x_j, \quad \sigma_{o_i} = \text{cov}(o_i, o_j),$$

and

$$\text{var} \left( \sum_{i=1}^{n} l_i x_i \right) = \sum_{i=1}^{n} \sigma^2_{l_i} x_i^2 + \sum_{j=1}^{n} \sum_{i=1, i \neq j}^{n} \sigma_{l_i l_j} x_i x_j, \quad \sigma_{l_i} = \text{cov}(l_i, l_j), \quad \forall i, j = 1, 2, \cdots, n.$$

Now, let us describe the portfolio selection model with interval values as follows.

**Model 5.1 Portfolio Selection Model with Interval Values**

Let $F_i \equiv (o_i, l_i)$, $\forall i = 1, 2, \cdots, n$, be interval values. The portfolio selection model
5.3 Fuzzy Statistical Test on the Portfolio Selection Model

with interval values is described as follows:

\[
\begin{align*}
\max & \sum_{i=1}^{n} E(o_i)x_i \\
\min & \sum_{i=1}^{n} E(l_i)x_i \\
s.t. & \sum_{i=1}^{n} \sqrt{\text{var}(o_i)} x_i \leq \varepsilon^2 \\
& \sum_{i=1}^{n} \sqrt{\text{var}(l_i)} x_i \leq \varepsilon^2 \\
& \sum_{i=1}^{n} x_i \leq 1 \\
& x_i \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{align*}
\]

(5.6)

where \( \varepsilon (\varepsilon \geq 0) \) represents the risk level, \( x^T V_o x = \sum_{i=1}^{n} \sigma_{o_i}^2 x_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sigma_{o_i o_j} x_i x_j \), \( \sigma_{o_i o_j} = \text{cov}(o_i, o_j) \), and \( x^T V_l x = \sum_{i=1}^{n} \sigma_{l_i}^2 x_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sigma_{l_i l_j} x_i x_j \), \( \sigma_{l_i l_j} = \text{cov}(l_i, l_j) \), \( \forall i, j = 1, 2, \ldots, n \).

To reduce the calculation load, we rewrite formula (5.6) into the following formula:

\[
\begin{align*}
\max & \sum_{i=1}^{n} E(o_i)x_i \\
\min & \sum_{i=1}^{n} E(l_i)x_i \\
s.t. & \sum_{i=1}^{n} \sqrt{\text{var}(o_i)} x_i = k \\
& \sum_{i=1}^{n} \sqrt{\text{var}(l_i)} x_i \leq k \\
& \sum_{i=1}^{n} x_i \leq 1 \\
& x_i \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{align*}
\]

(5.7)

Note that because

\[
x^T V_o x = \sum_{i=1}^{n} \sigma_{o_i}^2 x_i^2 + \sum_{j=1}^{n} \sum_{i=1, i \neq j}^{n} \sigma_{o_i o_j} x_i x_j \leq (\sum_{i=1}^{n} \sigma_{o_i} x_i)^2 \leq \varepsilon^2,
\]

we set \( (\sum_{i=1}^{n} \sigma_{o_i} x_i)^2 = k_1^2 \), i.e., \( \sum_{i=1}^{n} \sigma_{o_i} x_i = k_1 \), where \( \sigma_{o_i} = \sqrt{\text{var}(o_i)} \). We can obtain another restriction in the same way, i.e., \( \sum_{i=1}^{n} \sigma_{l_i} x_i = k_2 \), where \( \sigma_{l_i} = \sqrt{\text{var}(l_i)} \). We can choose a \( k \) such that \( \sum_{i=1}^{n} \sqrt{\text{var}(o_i)} x_i = k_1 = k \) and \( \sum_{i=1}^{n} \sqrt{\text{var}(l_i)} x_i = k_2 \leq k \).

Therefore, we can obtain the optimal solution of the model 5.1 in (5.7) by choosing different values for \( k (k \geq 0) \), which indicates the acceptable risk level. We obtain the optimal solution vector \( x^* = [x_1^*, x_2^*, \ldots, x_n^*]' \).
Moreover, when we obtain the optimal solution vector \( x^* = [x^*_1, x^*_2, \cdots, x^*_n]' \), we can calculate the fuzzy expected return \( E(R(x^*)) \):

\[
E[R(x^*)] \equiv \left( \sum_{i=1}^{n} E(o_i)x^*_i, \sum_{i=1}^{n} E(l_i)x^*_i \right).
\]

Now, we have defined the portfolio selection model with interval values. The model can have many solutions. The solutions depend on the risk level \( k \). Let us describe how to choose the risk level \( k \) in the following subsection.

### 5.3.2 Fuzzy Statistical Test for the Portfolio Selection Model

First, we define a \( T \)-test with interval data.

**Definition 5.4 \( T \)-test with interval values**

For interval values \( F_i = [a_i, b_i] \equiv (o_i, l_i), \) samples \( o_1, o_2, \cdots, o_n \) can be approximated as a normally distributed stochastic quantity \( O \sim N(m_o, \sigma^2_o) \) by Theorem 5.1, and the samples \( l_1, l_2, \cdots, l_n \) can be approximated as a normally distributed stochastic quantity \( L \sim N(m_l, \sigma^2_l) \) by Theorem 5.1. Hence, the bivariate normal distribution of the 2-dimensional random vector \( F \) can be written in the notation \( F \sim N(m, \Sigma) \), where \( m = (m_o, m_l) \) is a mean vector and \( \Sigma = \begin{bmatrix} \sigma^2_o & 0 \\ 0 & \sigma^2_l \end{bmatrix} \) because \( o \) and \( l \) are independent.

Statistics for each hypothesis are written as follows.

\[
H_0 : \quad m = m_0, \quad \text{for statistic vector } T \equiv (T_o, T_l),
\]

where \( m_0 \equiv (m_{o0}, m_{l0}) \) is a specified vector, \( T_o = \frac{\bar{X}_{o_n} - m_{o0}}{S_{o_n}/\sqrt{n}} \) and \( T_l = \frac{\bar{X}_{l_n} - m_{l0}}{S_{l_n}/\sqrt{n}} \).

Note that, here \( m_{o0} = \sum_{i=1}^{n} E(o_i)x^* \) and \( m_{l0} = \sum_{i=1}^{n} E(l_i)x^* \). We set \( m_0 = E[R(x^*)] \) in this chapter.

The acceptance region \( \mathcal{A} \) for \( T \) under probability \( \alpha \) for an error of the first type is defined as

\[
\mathcal{A} = \mathcal{A}_o \times \mathcal{A}_l.
\]

Note that

\[
\mathcal{A}_o = \{ t_o \in \mathbb{R} \mid t_o = |\frac{\bar{X}_{o_n} - m_{o0}}{S_{o_n}/\sqrt{n}}| \leq t_{n-1;1-0.5}, \} \quad \text{and}
\]

\[
5.6
\]
5.3 Fuzzy Statistical Test on the Portfolio Selection Model

\[ \mathcal{A}_l = \{ t_l \in \mathbb{R} \mid t_l = \left| \frac{\bar{X}_{l_n} - m_{l_0}}{S_{l_n}} \right| \leq t_{n-1;1-\frac{\alpha}{2}} \}, \]  

where \( \bar{X}_{o_n} \) is the sample mean of \( o \), \( S_{o_n} \) is the sample standard deviation of \( o \), \( \bar{X}_{l_n} \) is the sample mean of \( l \), \( S_{l_n} \) is the sample standard deviation of \( l \), \( n \) is the sample size and \( t_{n-1;1-\frac{\alpha}{2}} \) is the \((1 - \frac{\alpha}{2})\)-fractile of the \( t \)-distribution with \( n - 1 \) degrees of freedom.

Now, we give the decision of the portfolio selection model based on the fuzzy statistical test.

**Definition 5.5 Fuzzy Statistical Test on the Portfolio Selection Model with Interval Values**

According to Definition 5.4 and Formula (5.7) in Model 5.1, we say that if \( m_0 \in \mathcal{A} \), then we do not reject the null hypothesis \( H_0 \). In this situation, the decision of \( K \) is

\[ K = \{ k \mid k \geq 0 \text{ such that } m_0 \equiv (m_{o_0}, m_{l_0}) \in \mathcal{A}, \text{ we do not reject } H_0 \} \].

Note that \( m_0 \in \mathcal{A} \) means that \( m_{o_0} \in \mathcal{A}_o \) and \( m_{l_0} \in \mathcal{A}_l \). Hence, we obtain the solution of Model 5.1 with different \( k \) and get a set \( K \).

Now, we give the procedure of solving the portfolio selection model with interval values by a fuzzy statistical test in the following subsection.

5.3.3 Procedure of Solving Portfolio Selection Model with Interval Values

The procedure can be written in the following to solve portfolio selection model with interval values by a fuzzy statistical test:

1. **Step 1.** Collect the fuzzy data and calculate the return of each exchange currency with interval values.
2. **Step 2.** Compute \( o_i \) and \( l_i \), \( \forall i = 1, 2, \cdots, n \).
3. **Step 3.** Identify the underlying distribution by simulating \( o_i \) and \( l_i \).
4. **Step 4.** Calculate the parameters for the expected value and variance in model (5.7).
5. **Step 5.** Solve the optimization model (5.7) with different values \( k \). Stop solving the model when the optimal solution is indicated with only one asset and \( \sum_{i=1}^{n} x_i = 1 \). (i.e. \( x^* = (1, 0, 0, 0, 0)' \) or \((0, 1, 0, 0, 0)' \) or \((0, 0, 1, 0, 0)' \) or \((0, 0, 0, 1, 0)' \) or \((0, 0, 0, 0, 1)' \).
5. RISK ASSESSMENT OF A PORTFOLIO SELECTION MODEL BASED ON A FUZZY STATISTICAL TEST

we stop solving the model.)

*Step 6.* Calculate $K$ by Definition 5.5.

*Step 7.* Obtain the optimal vector solution $x^*$ with different $k$ in $K$.

*Step 8.* Compute the possibility distribution of the fuzzy expected return $R(x^*)$ with different $k$ in $K$.

To illustrate our proposed effective meanings and approaches to obtain efficient portfolios, we exemplify a real portfolio selection problem in the following section.

5.4 Empirical Studies

**Example 5.1** We selected five exchange currencies (USD, EUR, AUD, GBP and CHF) from the Bank of Tokyo-Mitsubishi. The original data came from the closing prices of every day from July 2010 to December 2010. There were 124 interval values in this period $[a, b]$, where $a$ is the minimum price, and $b$ is the maximum price in one day. We presented some interval values in Table 5.1.

<table>
<thead>
<tr>
<th></th>
<th>July 1, 2010</th>
<th>July 2, 2010</th>
<th>⋮</th>
<th>Dec. 31, 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>USD</strong></td>
<td>[86.94, 88.55]</td>
<td>[87.30, 88.20]</td>
<td>⋮</td>
<td>[81.26, 81.84]</td>
</tr>
<tr>
<td><strong>EUR</strong></td>
<td>[106.80,109.84]</td>
<td>[109.49,110.65]</td>
<td>⋮</td>
<td>[107.70,108.56]</td>
</tr>
<tr>
<td><strong>AUD</strong></td>
<td>[72.82, 73.99]</td>
<td>[73.69, 74.66]</td>
<td>⋮</td>
<td>[82.45, 83.14]</td>
</tr>
<tr>
<td><strong>GBP</strong></td>
<td>[130.26,133.06]</td>
<td>[132.89,133.61]</td>
<td>⋮</td>
<td>[125.27,126.67]</td>
</tr>
<tr>
<td><strong>CHF</strong></td>
<td>[81.20, 82.83]</td>
<td>[82.02, 82.90]</td>
<td>⋮</td>
<td>[86.17, 87.25]</td>
</tr>
</tbody>
</table>

Suppose that we buy the five exchange currencies with opening prices on July 1. We assume that an investor buy 5 exchange currencies on July 1, and do not take any action from the day he bought around half a year. Our objective is to choose a portfolio that maximizes the return (interval values) on the investment under some constraints on the selection with different risks $k$. Moreover, we make the decision for selecting the best return by a fuzzy statistical test.
First, we calculated the interval returns \([A, B]\) of each exchange currency, where \(A=a\)-opening price on July 1 and \(B=b\)-opening price on July 1. The original prices on July 1 of each exchange currency were 88.35, 107.15, 73.51, 130.91 and 81.36, respectively. We give the results in Table 5.2.

**Table 5.2: Interval Returns of Each Exchange Currency**

<table>
<thead>
<tr>
<th></th>
<th>([A, B])</th>
<th>([-7.08, -6.50])</th>
<th>([0.55, 1.41])</th>
<th>([8.94, 9.63])</th>
<th>([-0.64, 2.15])</th>
<th>([-5.63, 4.24])</th>
<th>([-0.15, 1.48])</th>
<th>([4.82, 5.90])</th>
</tr>
</thead>
<tbody>
<tr>
<td>USD</td>
<td>[-1.41, 0.21]</td>
<td>[-1.04, -0.14]</td>
<td>⋯</td>
<td>[-7.08, -6.50]</td>
<td>⋯</td>
<td>[0.55, 1.41]</td>
<td>⋯</td>
<td>[4.82, 5.90]</td>
</tr>
<tr>
<td>EUR</td>
<td>[-0.35, 2.69]</td>
<td>[2.35, 3.51]</td>
<td>⋯</td>
<td>[8.94, 9.63]</td>
<td>⋯</td>
<td>[0.55, 1.41]</td>
<td>⋯</td>
<td>[4.82, 5.90]</td>
</tr>
<tr>
<td>AUD</td>
<td>[-0.69, 0.48]</td>
<td>[0.18, 1.15]</td>
<td>⋯</td>
<td>[-0.64, 2.15]</td>
<td>⋯</td>
<td>[0.18, 1.15]</td>
<td>⋯</td>
<td>[0.18, 1.15]</td>
</tr>
<tr>
<td>GBP</td>
<td>[-0.64, 2.15]</td>
<td>[1.99, 2.70]</td>
<td>⋯</td>
<td>[-0.64, 2.15]</td>
<td>⋯</td>
<td>[1.99, 2.70]</td>
<td>⋯</td>
<td>[1.99, 2.70]</td>
</tr>
<tr>
<td>CHF</td>
<td>[-0.15, 1.48]</td>
<td>[0.66, 1.55]</td>
<td>⋯</td>
<td>[-0.15, 1.48]</td>
<td>⋯</td>
<td>[0.66, 1.55]</td>
<td>⋯</td>
<td>[0.66, 1.55]</td>
</tr>
</tbody>
</table>

Then, we calculated the central point \(o = \frac{A+B}{2}\) and radius \(l = \frac{B-A}{2}\). Table 5.3 presents the data.

**Table 5.3: Central Point \(o\) and Radius \(l\) of Each Interval Values \([A, B]\)**

<table>
<thead>
<tr>
<th></th>
<th>((o, l))</th>
<th>((-6.79, 0.29))</th>
<th>((0.98, 0.43))</th>
<th>((-4.93, 0.70))</th>
<th>((5.36, 0.54))</th>
</tr>
</thead>
<tbody>
<tr>
<td>USD</td>
<td>(-0.60, 0.81)</td>
<td>(-0.59, 0.45)</td>
<td>⋯</td>
<td>(-6.79, 0.29)</td>
<td>⋯</td>
</tr>
<tr>
<td>EUR</td>
<td>(1.17, 1.52)</td>
<td>(2.93, 0.58)</td>
<td>⋯</td>
<td>(0.98, 0.43)</td>
<td>⋯</td>
</tr>
<tr>
<td>AUD</td>
<td>(-0.11, 0.59)</td>
<td>(0.67, 0.49)</td>
<td>⋯</td>
<td>(9.29, 0.35)</td>
<td>⋯</td>
</tr>
<tr>
<td>GBP</td>
<td>(0.76, 1.40)</td>
<td>(2.35, 0.36)</td>
<td>⋯</td>
<td>(-4.93, 0.70)</td>
<td>⋯</td>
</tr>
<tr>
<td>CHF</td>
<td>(0.66, 0.81)</td>
<td>(1.11, 0.44)</td>
<td>⋯</td>
<td>(5.36, 0.54)</td>
<td>⋯</td>
</tr>
</tbody>
</table>

We simulated the values \(o\) and \(l\), respectively. We obtain the probability distributions \(O\) and \(L\) for each respective exchange currency. Table 5.4 presents the results.

**Table 5.4: Interval Probabilities of Each Exchange Currency**

<table>
<thead>
<tr>
<th></th>
<th>(\text{LOG})</th>
<th>(\text{W})</th>
<th>(\text{N})</th>
<th>(\Gamma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USD</td>
<td>((-6.79, 0.29))</td>
<td>((-6.79, 0.29))</td>
<td>((-6.79, 0.29))</td>
<td>((-6.79, 0.29))</td>
</tr>
<tr>
<td>EUR</td>
<td>((0.98, 0.43))</td>
<td>((0.98, 0.43))</td>
<td>((0.98, 0.43))</td>
<td>((0.98, 0.43))</td>
</tr>
<tr>
<td>AUD</td>
<td>((-4.93, 0.70))</td>
<td>((-4.93, 0.70))</td>
<td>((-4.93, 0.70))</td>
<td>((-4.93, 0.70))</td>
</tr>
<tr>
<td>GBP</td>
<td>((5.36, 0.54))</td>
<td>((5.36, 0.54))</td>
<td>((5.36, 0.54))</td>
<td>((5.36, 0.54))</td>
</tr>
</tbody>
</table>

Note that the abbreviation \(\text{LOG}\) denotes logistic distribution, \(\text{W}\), \(\text{N}\), and \(\Gamma\) denote Weibull distribution, normal distribution and gamma distribution, respectively.

When we knew the distribution function, we used the moment method estimator (MME) to estimate the parameter for each distribution function. Hence, we could find
5. RISK ASSESSMENT OF A PORTFOLIO SELECTION MODEL
BASED ON A FUZZY STATISTICAL TEST

Table 5.4: Parameters of Probability Distribution Functions for Interval Values

<table>
<thead>
<tr>
<th></th>
<th>O</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>USD</td>
<td>LOG(−4.23, 1.21)</td>
<td>W(0.45, 3.26)</td>
</tr>
<tr>
<td>EUR</td>
<td>LOG( 4.44, 1.33)</td>
<td>Γ (7.69, 0.08)</td>
</tr>
<tr>
<td>AUD</td>
<td>LOG( 6.03, 1.56)</td>
<td>Γ (7.51, 0.07)</td>
</tr>
<tr>
<td>GBP</td>
<td>N ( 0.80, 2.41²)</td>
<td>Γ (9.87, 0.06)</td>
</tr>
<tr>
<td>CHF</td>
<td>W ( 3.01, 1.58)</td>
<td>Γ (9.99, 0.05)</td>
</tr>
</tbody>
</table>

out the expected values and variances by using those parameters. Table 5.5 shows the results.

Table 5.5: Expected Values and Variances for Interval Values

<table>
<thead>
<tr>
<th></th>
<th>O₁</th>
<th>O₂</th>
<th>O₃</th>
<th>O₄</th>
<th>O₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected value</td>
<td>-4.23</td>
<td>4.44</td>
<td>6.03</td>
<td>0.80</td>
<td>2.70</td>
</tr>
<tr>
<td>Variance</td>
<td>4.79</td>
<td>5.78</td>
<td>8.03</td>
<td>2.41²</td>
<td>3.06</td>
</tr>
<tr>
<td>L₁</td>
<td>L₂</td>
<td>L₃</td>
<td>L₄</td>
<td>L₅</td>
<td></td>
</tr>
<tr>
<td>Expected value</td>
<td>0.40</td>
<td>0.62</td>
<td>0.53</td>
<td>0.59</td>
<td>0.50</td>
</tr>
<tr>
<td>Variance</td>
<td>0.02</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Now, we have all the data that we need in our portfolio selection model with interval values. We put these data in Model (5.7) and present the results as follows:

\[
\begin{align*}
\max & \quad -4.23x₁ + 4.44x₂ + 6.03x₃ + 0.80x₄ + 2.70x₅ \\
\min & \quad 0.40x₁ + 0.62x₂ + 0.53x₃ + 0.59x₄ + 0.50x₅ \\
s.t. & \quad \sqrt{4.79}x₁ + \sqrt{5.78}x₂ + \sqrt{8.03}x₃ + \sqrt{5.81}x₄ \\
& \quad + \sqrt{3.06}x₅ = k \\
& \quad \sqrt{0.02}x₁ + \sqrt{0.05}x₂ + \sqrt{0.04}x₃ + \sqrt{0.04}x₄ \\
& \quad + \sqrt{0.02}x₅ \leq k \\
& \quad x₁ + x₂ + x₃ + x₄ + x₅ \leq 1 \\
& \quad xᵢ \geq 0 \quad \forall i = 1, 2, \cdots, n
\end{align*}
\]

(5.12)
5.4 Empirical Studies

We rewrite the model with estimated parameters as follows:

\[
\begin{align*}
\max & \quad -4.23x_1 + 4.44x_2 + 6.03x_3 + 0.80x_4 + 2.70x_5 \\
\min & \quad 0.40x_1 + 0.62x_2 + 0.53x_3 + 0.59x_4 + 0.50x_5 \\
\text{s.t.} & \quad 2.19x_1 + 2.40x_2 + 2.83x_3 + 2.41x_4 + 1.75x_5 = k \\
& \quad 0.14x_1 + 0.22x_2 + 0.20x_3 + 0.20x_4 + 0.14x_5 \leq k \\
& \quad x_1 + x_2 + x_3 + x_4 + x_5 \leq 1 \\
& \quad x_i \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{align*}
\]  

We solved Model (5.13) by using GP-IGP (Linear and Integer Goal Programming). The result depends on the selection with different values of \( k \). We gave the value \( k \) greater than zero and accurate to second decimal places in this example. We present the result in Table 5.6.

**Table 5.6:** Fuzzy Statistical Test for the Results of Model (5.13) with Different Conditions \( k \)

<table>
<thead>
<tr>
<th>( k )</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>2.5</th>
<th>2.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sum_{i=1}^{n} x_i )</td>
<td>0.23</td>
<td>0.46</td>
<td>0.68</td>
<td>0.91</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x^* )</td>
<td>[23.0, 0, 0, 0, 0]'</td>
<td>[46.0, 0, 0, 0, 0]'</td>
<td>[68.0, 0, 0, 0, 0]'</td>
<td>[91.0, 0, 0, 0, 0]'</td>
<td>[52.0, 48, 0, 0, 0]'</td>
<td>[36.0, 64, 0, 0, 0]'</td>
</tr>
<tr>
<td>( m_0 )</td>
<td>(-0.97, 0.09)</td>
<td>(-1.93, 0.18)</td>
<td>(-2.90, 0.27)</td>
<td>(-3.86, 0.37)</td>
<td>(0.74, 0.40)</td>
<td>(2.34, 0.48)</td>
</tr>
<tr>
<td>( m )</td>
<td>(-0.95, 0.09)</td>
<td>(-1.90, 0.18)</td>
<td>(-2.81, 0.28)</td>
<td>(-3.77, 0.37)</td>
<td>(0.68, 0.45)</td>
<td>(2.29, 0.47)</td>
</tr>
<tr>
<td>( S_m )</td>
<td>(0.47, 0.03)</td>
<td>(0.95, 0.07)</td>
<td>(1.41, 0.11)</td>
<td>(1.89, 0.15)</td>
<td>(1.05, 0.13)</td>
<td>(1.36, 0.14)</td>
</tr>
<tr>
<td>( A )</td>
<td>[-1.03, -0.86]</td>
<td>[-2.07, -1.73]</td>
<td>[-3.06, -2.56]</td>
<td>[-4.10, -3.43]</td>
<td>[-0.97, -0.86]</td>
<td>[-2.05, 2.54]</td>
</tr>
<tr>
<td>( \times [0.08, 0.10] )</td>
<td>( \times [0.17, 0.20] )</td>
<td>( \times [0.25, 0.30] )</td>
<td>( \times [0.34, 0.40] )</td>
<td>( \times [0.43, 0.48] )</td>
<td>( \times [0.44, 0.49] )</td>
<td></td>
</tr>
</tbody>
</table>

Decision: \( m_0 \in \mathcal{A} \)

<table>
<thead>
<tr>
<th>( k )</th>
<th>2.7</th>
<th>2.8</th>
<th>2.81</th>
<th>2.82</th>
<th>2.83</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sum_{i=1}^{n} x_i )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x^* )</td>
<td>[2.0, 0, 8, 0, 0]'</td>
<td>[65.0, 95, 0, 0, 0]'</td>
<td>[93.0, 97, 0, 0, 0]'</td>
<td>[92.0, 98, 0, 0, 0]'</td>
<td>[0.0, 1, 0, 0, 0]'</td>
</tr>
<tr>
<td>( m_0 )</td>
<td>(3.95, 0.50)</td>
<td>(5.55, 0.52)</td>
<td>(7.51, 0.53)</td>
<td>(7.51, 0.53)</td>
<td>(6.03, 0.53)</td>
</tr>
<tr>
<td>( m )</td>
<td>(3.90, 0.48)</td>
<td>(5.41, 0.50)</td>
<td>(7.41, 0.50)</td>
<td>(5.71, 0.50)</td>
<td>(5.92, 0.50)</td>
</tr>
<tr>
<td>( S_m )</td>
<td>(1.87, 0.15)</td>
<td>(2.42, 0.17)</td>
<td>(2.90, 0.18)</td>
<td>(2.53, 0.18)</td>
<td>(2.61, 0.18)</td>
</tr>
<tr>
<td>( A )</td>
<td>[3.57, 4.24]</td>
<td>[4.98, 5.84]</td>
<td>[5.17, 6.06]</td>
<td>[5.26, 6.17]</td>
<td>[5.45, 6.38]</td>
</tr>
<tr>
<td>( \times [0.46, 0.51] )</td>
<td>( \times [0.47, 0.53] )</td>
<td>( \times [0.47, 0.53] )</td>
<td>( \times [0.47, 0.53] )</td>
<td>( \times [0.47, 0.54] )</td>
<td></td>
</tr>
</tbody>
</table>

Decision: \( m_0 \in \mathcal{A} \)

**Explanation of decision with \( T \)-test**

In Table 5.6, for example, when \( k = 2.5 \), we have \( x^* = [0.52, 0, 0.48, 0, 0]' \). We calculated the return

\[
R = \sum_{i=1}^{n} r_i x_i = r_1 x_1 + r_2 x_2 + r_3 x_3 + r_4 x_4 + r_5 x_5
\]

\[
= 0.52 \times r_1 + 0.48 \times r_3.
\]

Therefore, we obtained 124 new data points.
5. RISK ASSESSMENT OF A PORTFOLIO SELECTION MODEL BASED ON A FUZZY STATISTICAL TEST

We calculated the expected value \( m \) and standard deviation \( S_n \) by Minitab15. The results are \( m \equiv (0.68, 0.45) \) and \( S_n \equiv (1.05, 0.13) \). The hypothesis was \( H_0 : m = (0.74, 0.46) \), for statistic vector \( T \equiv (T_o, T_l) \), where \( T_o = \left| \frac{0.68 - 0.74}{1.05} \right| = 0.55 \) and \( T_l = \left| \frac{0.45 - 0.46}{0.13} \right| = 0.17 \). The 95 percent confidence interval was \([0.49, 0.87]\) and \([0.43, 0.48]\), respectively.

Because \((0.74, 0.46) \in [0.49, 0.87] \times [0.43, 0.48]\), we accepted the hypothesis. Note that we say \((0.74, 0.46) \in [0.49, 0.87] \times [0.43, 0.48]\); this means that \(0.74 \in [0.49, 0.87]\) and \(0.46 \in [0.43, 0.48]\).

Hence, we obtained the fuzzy expected return

\[
E[R(x^*)] = (\sum_{i=1}^{n} E(o_i)x_i^*) + (\sum_{i=1}^{n} E(l_i)x_i^*) = (0.74, 0.46).
\]

The interval value of the fuzzy expected return was \([0.28, 1.20]\).

Using the same method for the other values of \( k \), we can also obtain the respective interval value of the fuzzy expected return. We present the results in Table 5.7.

<table>
<thead>
<tr>
<th>( k )</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>2.5</th>
<th>2.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^* )</td>
<td>([0.23, 0.6, 0.0, 0])</td>
<td>([0.46, 0.6, 0, 0])</td>
<td>([0.8, 0.6, 0, 0])</td>
<td>([0.91, 0.6, 0, 0])</td>
<td>([0.52, 0.48, 0, 0])</td>
<td>([0.36, 0.64, 0, 0])</td>
</tr>
<tr>
<td>( E[R(x^*)] )</td>
<td>([-0.97, 0.99])</td>
<td>([-1.93, 0.18])</td>
<td>([-2.96, 0.37])</td>
<td>([-3.86, 0.37])</td>
<td>([-0.74, 0.46])</td>
<td>([2.34, 0.48])</td>
</tr>
<tr>
<td>Interval Value</td>
<td>([-1.06, -0.88])</td>
<td>([-2.11, -1.75])</td>
<td>([-3.17, -2.63])</td>
<td>([-4.23, -3.49])</td>
<td>([-0.28, 1.20])</td>
<td>([1.86, 2.82])</td>
</tr>
</tbody>
</table>

In Table 5.7, we can see that we accepted the hypothesis for \( k \leq 2.83 \) and that we had a stable return for \( k \geq 2.83 \) because we had the same expected return of radius. Moreover, we cannot solve model (5.13) for \( k > 2.83 \). We obtained a negative return for \( k \leq 2 \) and the maximum return for \( k = 2.83 \). We got the optimal solution \( x^* = (0, 0, 1, 0, 0) \) when \( k = 2.83 \). Therefore, we stop solving the model.

In this example, we conclude that the maximum fuzzy expected return was \([5.50, 6.56]\). We present a scatterplot of \( m_0 = E[R(x^*)] \equiv (m_{o0}, m_{l0}) \) with respect to different values of \( k \) in Figure 5.1.
Example 5.2 In Reference (85), Zhang implemented the concept of the \( \gamma \)-level to deal with the optimization model. He also adopted an additional condition of the upper and lower bounds in the model proposed by Markowitz. We applied his model to solve our problem but deleted the upper bound and lower bound of his example.

First, we calculated the expected value of 124 interval returns by Definition 2.5. We obtained 5 interval numbers as follows: USD=\( r_1 = [-4.56, -3.73] \), EUR=\( r_2 = [3.59, 4.89] \), AUD=\( r_3 = [5.41, 6.43] \), GBP=\( r_4 = [0.17, 1.44] \), and CHF=\( r_5 = [2.05, 3.06] \).

Hence, the lower possibilistic mean-standard deviation model was

\[
\begin{align*}
\max & \quad -4.56x_1 + 3.59x_2 + 5.41x_3 + 0.17x_4 + 2.05x_5 \\
\text{s.t.} & \quad x_1 + x_2 + x_3 + x_4 + x_5 \leq 1 \\
& \quad x_i \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{align*}
\]

and the upper possibilistic mean-standard deviation model was

\[
\begin{align*}
\max & \quad -3.73x_1 + 4.89x_2 + 6.43x_3 + 1.44x_4 + 3.06x_5 \\
\text{s.t.} & \quad x_1 + x_2 + x_3 + x_4 + x_5 \leq 1 \\
& \quad x_i \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{align*}
\]
5. RISK ASSESSMENT OF A PORTFOLIO SELECTION MODEL BASED ON A FUZZY STATISTICAL TEST

Table 5.8: Fuzzy Statistical Test for The Results of Model (5.15) and (5.16) with Different Conditions

<table>
<thead>
<tr>
<th>$\sum_{i=1}^{5} x_i$</th>
<th>0.23</th>
<th>0.46</th>
<th>0.68</th>
<th>0.91</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^*$</td>
<td>[0, 0, 0.23, 0, 0]$^\prime$</td>
<td>[0, 0, 0.46, 0, 0]$^\prime$</td>
<td>[0, 0, 0.68, 0, 0]$^\prime$</td>
<td>[0, 0, 0.91, 0, 0]$^\prime$</td>
<td>[0, 0, 1, 0, 0]$^\prime$</td>
</tr>
<tr>
<td>$[L^<em>, U^</em>]$</td>
<td>[1.24, 1.48]</td>
<td>[2.49, 2.96]</td>
<td>[3.68, 4.37]</td>
<td>[4.92, 5.85]</td>
<td>[5.41, 6.43]</td>
</tr>
<tr>
<td>$m_0$</td>
<td>(1.36, 0.12)</td>
<td>(2.725, 0.235)</td>
<td>(4.025, 0.345)</td>
<td>(5.385, 0.465)</td>
<td>(5.92, 0.51)</td>
</tr>
<tr>
<td>$m$</td>
<td>(1.36, 0.12)</td>
<td>(2.72, 0.23)</td>
<td>(4.02, 0.35)</td>
<td>(5.39, 0.46)</td>
<td>(5.92, 0.51)</td>
</tr>
<tr>
<td>$S_n$</td>
<td>(0.60, 0.04)</td>
<td>(1.20, 0.09)</td>
<td>(1.78, 0.13)</td>
<td>(2.38, 0.17)</td>
<td>(2.61, 0.19)</td>
</tr>
<tr>
<td>$A$</td>
<td>[1.26, 1.47]</td>
<td>[2.51, 2.94]</td>
<td>[3.71, 4.34]</td>
<td>[4.97, 5.81]</td>
<td>[5.46, 6.38]</td>
</tr>
<tr>
<td>$\times [0.11, 0.12]$</td>
<td>$\times [0.22, 0.25]$</td>
<td>$\times [0.32, 0.37]$</td>
<td>$\times [0.43, 0.49]$</td>
<td>$\times [0.47, 0.54]$</td>
<td></td>
</tr>
</tbody>
</table>

Decision $m_0 \in A$, $m_0 \in A$, $m_0 \in A$, $m_0 \in A$, $m_0 \in A$.

Table 5.8 shows the results, where $L^*$ denotes the optimal solution in model (5.15), and $U^*$ denotes as the optimal solution in model (5.16).

From Table 5.8, we can see that when we choose $\sum_{i=1}^{5} x_i = 1$, the optimal solution of model (5.15) and (5.16) are the same as $x^* = [0, 0, 1, 0, 0]^\prime$.

Hence, we obtained the fuzzy expected return

$$E[R(x^*)] \equiv \left( \sum_{i=1}^{n} E(o_i)x_i^*, \sum_{i=1}^{n} E(l_i)x_i^* \right)$$

$$= (5.92, 0.51). \quad (5.17)$$

The interval value of the fuzzy expected return was $[5.41, 6.43]$.

5.5 Summary of Chapter 5

5.5.1 Discussion

In this chapter, we attempted to establish a fuzzy statistical test. We proposed a method to defuzzify fuzzy data. That is, we used central point and radius instead of interval data. Therefore, the central point and radius were simplified to real numbers and had statistic characteristic. We estimated the probability distribution by using central point and radius and calculated the expected value and variance based on the estimated parameters of the underlying probability distribution. We supported the efficacy of the proposed method through an application of maximizing investment.
portfolio of foreign exchange currencies. An empirical study of a portfolio selection model was conducted based on a fuzzy statistical test in Example 5.1.

Example 5.1 showed that we chose only one exchange currency (AUD) and had the maximum expected return in $(5.50, 6.56)$ when $k = 2.83$. We say that we have a stable return for $k \geq 2.81$ because we have the same expected return of radius. Moreover, we get a negative value of return when the value $k$ is less than or equal to 2. We can see that we accepted all the values $k$ for selecting the best return in Example 5.1. That is, we chose the expected return when $k = 0.5, 1, 1.5, 2, 2.5, 2.6, 2.7, 2.8, 2.81, 2.82, 2.83$. But in fact we do not accept all the expected return because we need to consider financial reports, experts’ individual experiences and other factors in real world. For example, we do not want to buy a negative expected return when the value $k$ is less than or equal to 2. Hence, in this example, we thought that an investor can consider to buy a portfolio when the value $k$ is greater than 2.

In Example 5.1, we took the same interval returns as Example 5.1 and used Zhang’s model (85) to solve the portfolio selection problem. The author took the left and right points of the interval returns to build separated models, thus obtaining the maximum expected return by solving these separated models. The results showed that Zhang obtained the maximum expected return by choosing only one exchange currency (AUD). (i.e. $x^* = (0, 0, 0.23, 0, 0)'$, or $(0, 0, 0.46, 0, 0)'$, or $(0, 0, 0.68, 0, 0)'$, or $(0, 0, 0.91, 0, 0)'$, or $(0, 0, 1, 0, 0)'$.) We made the value $\sum_{i=1}^{5} x_i$ equal to 0.23, 0.46, 0.68, 0.91 and 1 in Table 5.8 to illustrate the optimal solutions. The value $\sum_{i=1}^{5} x_i$ in Table 5.8 is the same as in Table 5.6. We made a decision by a fuzzy statistical test in this example and accepted all the solution $(L^*, U^*)$ under different conditions. We could see that we got positive values of expected return in different conditions in Table 5.8. The expected return is better than the expected returns in Example 5.1 when the value $\sum_{i=1}^{5} x_i$ is equal to 0.23, 0.46, 0.68 and 0.91, but will get less return when $\sum_{i=1}^{5} x_i = 1$ and choose only one exchange currency (AUD). (i.e. the optimal solution is $x^* = (0, 0, 1, 0, 0)'$.) We conclude that although we accepted all the expected return in this example, but it is shortcoming that the author did not consider the risk level and only chose one exchange currency in different conditions.

In both these two examples, we can see that our model can give a greater expected return than the expected return in Zhang’s model when $k = 2.83$. Moreover, we provide different risk levels for investors to make decision. The fuzzy statistical test...
also provides reasonable results in our model. We not only make a decision by a fuzzy statistical test but also consider financial reports, experts’ individual experiences and other factors in real world.

The results base on a fuzzy statistical test can indicate two informations as follows.

1. In this chapter, we tested the expected return by a fuzzy statistical test, the results indicated whether we should accept or reject the expected return.

2. Since the expected return was solved from the portfolio selection model and the parameters in the model were calculated by estimated parameters of underlying distribution function, we conclude that if we accept the hypothesis (i.e. expected return), it means that it is no problem from data extraction to get an expected return based on the value $k$.

In this chapter, we provide the risk level $k$ for investors to make decision. We need to decide the value $k$ first and solve the linear programming model many times until we get the solution with only one exchange currency. Because of setting the value $k$ in the model, we have many expected returns which depend on the value $k$. We obtain a maximum return with different risk levels in our model and make a decision for selecting the best return by a fuzzy statistical test. We conclude that it is conservative investment and more objective for investors to make decision when they buy many exchange currencies. We also conclude that the evaluation by the fuzzy statistical test enables us to obtain a stable expected return and low risk investment with different choices based on the risk level $k$.

5.5.2 Conclusions

In this chapter, we established a statistical test of fuzzy data which is called as fuzzy statistical test. We introduced a concept for "defuzzifying" fuzzy data into real numbers. That is, we use the central point and radius instead of the interval data. Hence, the central point and radius will have the statistic characteristics as mean and variance, and the conventional statistic test can be applied. In order to illustrate the efficacy of the proposed method, we introduced an application of maximizing investment portfolio of foreign exchange currencies.

The portfolio selection model was built by using expected value and variance of central point and radius. The expected value and variance was calculated by the
estimated parameters of underlying distribution function. We evaluated the best return by a fuzzy statistical test. In this procedure, from data extraction to fuzzy statistical test, it is no doubt that the model can deal with the interval data, so does the fuzzy statistical test, because we had ”defuzzifying” fuzzy data into real numbers before we solved the portfolio selection model. Hence, the model becomes to a traditional linear programming model.

The empirical studies showed that we could provide many risk levels and expected returns. It’s more objective for investors to make decision when they buy exchange currencies. But we still have further points need to improve in the future as follows:

1. In this chapter, we just considered the interval returns. We thought that if we can estimate the returns with triangular fuzzy numbers or trapezoid fuzzy numbers in the future, it will make the proposed method more realizable.

2. In the proposed portfolio selection model, we gave a constraint inequality with risk level \( k \) which was given in advance. We made the value \( k \) greater than zero and accurate to second decimal places in this chapter. We thought that we can give more risk levels and results for investors to make decision in the future.

3. In fact that the financial market is affected by many non-probabilistic factors and the future returns of risky assets cannot be predicted accurately in any uncertain economic environment. Although we can evaluate and select the best return by a fuzzy statistical test, we thought that we also need to consider financial reports, experts’ individual experiences and other factors in real world.
6

A Parametric Assessment Approach to Solving Facility Location Problems with Fuzzy Demands

6.1 Introduction and Literature Review

Facility location selection is one of the most critical and strategic issues in supply-chain design and management. Choosing proper locations for facilities or even selecting the set of location alternatives is often one of difficult elements in planning process. Especially, it is complicated to optimize the locations because of having many constraints such as the developments of the city, the environment where we cite the facility, and transportation in that city.

Many research works have dealt with facility location problems to find the most effective and efficient method. Larson and Sadig (34) discussed facility location problem with $L_1$ metric in the presence of barriers to travel. Moreover, Katz and Cooper (24) discussed $L_2$ distance facility location problem with only one forbidden circle. In both of their studies, uncertain situations are not considered in facility-location problems. But sometimes randomness and fuzziness may coexist in real facility-location problems (43). We face vague information when describing data in natural language (50). Hence, this chapter deals with uncertainty of both randomness and fuzziness.
Using fuzzy random variables initiated by Kwakernaak (29)(30) is one of appropriate ways to describe such hybrid uncertainty. Wang, et al. (73) have developed a two-stage fuzzy-random facility-location model with recourse under a hybrid uncertain environment. Other approaches to the treatment of randomness and fuzziness may refer Wang and Watada (72), Puri and Ralescu (57)(58), Klement, et al. (25), Kruse (27), Kruse and Meyer (28), and Negoita and Ralescu (49).

In this chapter, we consider the uncertainty of demands in facility location problems, where uncertain demand is called as fuzzy demand.

Using fuzzy demand, we build up a facility location model. To determine the parameters of fuzzy demand, in surveying the demand of customers we use fuzzy data expression. Hence, our objective is to solve the facility location model. In the literature, we can find various discussions on the parameter of fuzzy demand. Some research works have discussed the fuzzy demand.

Fung, et al. (12) used uncertain market demands and capacities in production environment. They developed a fuzzy multiproduction model and solved it by using parametric programming. Also Ohtake and Nishida (52) proposed a parametric facility location problem (PFLP) based on various demands of customers and gave the new concept of a parametric analysis to improve the computational efficiency of the algorithm. Moreover, Nishida, et al. (51) employed the uncertain demand in the facility location problem and proposed a branch-and-bound algorithm to find an exact solution. Additionally, Tohyama, et al. (67) treated an uncapacitated facility location problem (UFLP) and proposed a genetic algorithm for solving UFLP. UFLP is a fundamental optimization problem to select of locations where some facilities supply the same service.

To express fuzzy random events in facility-location problems, the first step is to understand the probability distribution function of fuzzy data (23). Conventional research works in the past did not recognize the underlying probability distribution function of fuzzy data in their problems. Lin, et al. (39) proposed a way to find out the probability distribution function of fuzzy demand and solved the facility location model by using two-stage fuzzy-random facility-location model with recourse. The method is more accurate in considering the fuzzy data of actual demand, but consumes much computation time in solving mathematical programming problem. In fact when we work with fuzzy data, the underlying probability distribution function of fuzzy data
is not known. On the other hand, it is not easy without recognizing any probability distribution function to describe vague data in statistical terms. Therefore, we must establish some statistical techniques to handle vague information and knowledge.

Hence, we propose in this chapter to find out the probability distribution function first and calculate the estimated expected value (EE value) by using the estimated parameters of underlying probability distribution functions in our facility location model. Moreover, we will define a defuzzification formula in this chapter.

Following Zadeh (83)(84), we will use fuzzy set theory and take the concept of fuzzy statistics into consideration. Recently, fuzzy statistical analysis and its applications are widely studied and applied to the social sciences (78)(79), which exhibit fuzzy statistic methods by using a central point. To deal with problems, almost all research studies used only the central points of fuzzy numbers. Lin, et al. (38) have proposed a method to find out the underlying probability distribution function of fuzzy data by using central point and radius. We will adopt this approach to find the underlying distribution function of fuzzy data and define fuzzy demand in our facility location model.

To solve the facility location model, we propose defuzzification formula to defuzzify the fuzzy demand, which is called the realization of the fuzzy demand (RFD). Lin, et al. (40) have defined a new weight function of central point and radius. The function is used to defuzzify the fuzzy data into a real number, but they did not assess their error. Hence, we will restructure the weight function to define the RFD formula by using the EE value of central point and radius in this chapter. The RFD formula combines the upper bound of $RFD (RFD^+)$ and the lower bound of $RFD (RFD^-)$. Moreover, an error assessment, which is called mean absolute percentage error of the fuzzy demand (MAPE-FD), is proposed in the $RFD$ formula. We also provide an empirical study to illustrate the proposed method in real-life location problem.

This chapter is organized as follows. Section 6.2 gives preliminary definitions we will use in this chapter. Section 6.3 gives some notations and states the location problems. Section 6.4 provides an empirical study of real-life location problem. Finally, we present comparisons and Conclusions in Section 6.5. We also give some future works in this section.
6.2 Preliminary Definitions

Let us introduce some definitions of LR-type fuzzy numbers (11) as follows.

**Definition 6.1** Let \( A = [a, b, c] \) be a triangular fuzzy number. The left width is \( L_1 = b - a \) and the right width is \( L_2 = c - b \). Its membership function is written in the following form:

\[
\mu_A(x) = \begin{cases} 
1 - \frac{b - x}{L_1}, & \text{if } b - L_1 \leq x \leq b, \\
1, & \text{if } x = b, \\
1 - \frac{x - b}{L_2}, & \text{if } b \leq x \leq b + L_2, \\
0, & \text{otherwise},
\end{cases}
\]

and it is also denoted by \( A \equiv (b, L_1, L_2) \).

Moreover, two operations on LR-type fuzzy numbers, namely fuzzy addition and scalar multiplications, are given as follows.

**Lemma 6.1** Let \( A_1 = (a_1, b_1, c_1) \) and \( A_2 = (a_2, b_2, c_2) \) be any two fuzzy numbers and \( \lambda \) be any real number. Then

(a) \( A_1 + A_2 = (a_1 + a_2, b_1 + b_2, c_1 + c_2) \).
(b) \( \lambda A_1 = (\lambda a_1, \lambda b_1, \lambda c_1), \forall \lambda \geq 0 \).

**Proof:** We give a brief proof as follows.

As a definition of fuzzy number, we adopt a \( \alpha \)-cut set (4). We use the notation \([A]_\alpha = [A^-, A^+]_\alpha\) to denote explicitly the \(\alpha\)-cuts of \( A \), \( \forall \alpha \in [0, 1] \).

Let \( \mu_{A_1}(x_1) \) be a membership function of \( A_1 \), \( \mu_{A_2}(x_2) \) be a membership function of \( A_2 \) and \( \mu_{A_1 + A_2}(x) \) be a membership function of \( A_1 + A_2 \).

For triangular fuzzy numbers \( A_1 = (a_1, b_1, c_1) \) and \( A_2 = (a_2, b_2, c_2) \), their \( \alpha \)-cuts are

\[
[A_1]_\alpha = \{x_1 \in \mathbb{R} : \mu_{A_1}(x_1) \geq \alpha \} = [A^-_1, A^+_1]_\alpha = [a_1 + \alpha(b_1 - a_1), c_1 - \alpha(c_1 - b_1)]_\alpha,
\]

and

\[
[A_2]_\alpha = \{x_2 \in \mathbb{R} : \mu_{A_2}(x_2) \geq \alpha \} = [A^-_2, A^+_2]_\alpha = [a_2 + \alpha(b_2 - a_2), c_2 - \alpha(c_2 - b_2)]_\alpha.
\]

It means that

\[ a_1 + \alpha(b_1 - a_1) \leq x_1 \leq c_1 - \alpha(c_1 - b_1), \]

71
6. A PARAMETRIC ASSESSMENT APPROACH TO SOLVING FACILITY LOCATION PROBLEMS WITH FUZZY DEMANDS

and

\[ a_2 + \alpha(b_2 - a_2) \leq x_2 \leq c_2 - \alpha(c_2 - b_2). \]

Hence, we have \((a_1 + a_2) + \alpha[(b_1 + b_2) - (a_1 + a_2)] \leq x_1 + x_2 \leq (c_1 + c_2) - \alpha[(c_1 + c_2) - (b_1 + b_2)].\)

That is

\[
[A_1 + A_2]_\alpha = \{ x = x_1 + x_2 \in \mathbb{R} : \mu_{A_1 + A_2}(x) \geq \alpha \} = [(a_1 + a_2) + \alpha((b_1 + b_2) - (a_1 + a_2)), (c_1 + c_2) - \alpha((c_1 + c_2) - (b_1 + b_2))]_\alpha.
\]

It means that \(A_1 + A_2 = (a_1 + a_2, b_1 + b_2, c_1 + c_2).\)

We have proved the part of (a).

In part of (b), since

\[
[A_1]_\alpha = \{ x_1 \in \mathbb{R} : \mu_{A_1}(x_1) \geq \alpha \} = [A_1^-, A_1^+]_\alpha = [a_1 + \alpha(b_1 - a_1), c_1 - \alpha(c_1 - b_1)]_\alpha.
\]

It means that

\[ a_1 + \alpha(b_1 - a_1) \leq x_1 \leq c_1 - \alpha(c_1 - b_1), \]

and \(\forall \lambda \geq 0,\)

\[ \lambda a_1 + \alpha(\lambda b_1 - \lambda a_1) \leq \lambda x_1 \leq \lambda c_1 - \alpha(\lambda c_1 - \lambda b_1). \]

That is

\[
[\lambda A_1]_\alpha = \{ \lambda x_1 \in \mathbb{R} : \mu_{\lambda A_1}(\lambda x_1) \geq \alpha \} = [\lambda a_1 + \alpha(\lambda b_1 - \lambda a_1), \lambda c_1 - \alpha(\lambda c_1 - \lambda b_1)]_\alpha.
\]

i.e. \(\lambda A_1 = (\lambda a_1, \lambda b_1, \lambda c_1), \forall \lambda \geq 0.\)

We have proved the part of (b).

Let us use Definition 6.1 and Lemma 6.1 to calculate the fuzzy expected value as Lemma 6.2.

**Lemma 6.2 Fuzzy Expected Value**

Let \(A = [a, b, c]\) be a triangular fuzzy numbers on a probability space \((\Omega, F, P)\), where \(\Omega\) is a sample space, \(F\) is a sigma-algebra, and \(P\) is a probability measure. Using
6.2 Preliminary Definitions

the left width $L_1$ and right width $L_2$, let us denote $A$ as $A \equiv (b, L_1, L_2)$. Then, the fuzzy expected value of $A$ is defined as follows

$$E(A) \equiv (E(b), E(L_1), E(L_2)).$$

**Proof**: We adopt three parameters, $b$, $L_1$ and $L_2$ to express a triangular fuzzy number $A$. Hence, $(b, L_1, L_2)$ is a random vector, whose components are random variables of scalar value on the same probability space $(\Omega, F, P)$.

Therefore, the expected value of a random vector $A$ is a fixed vector $E(A)$, whose elements are the expected values of the respective random variables.

That is

$$E(A) \equiv (E(b), E(L_1), E(L_2)).$$

To find out the realization of fuzzy data, Lin, *et al.* (40) have defined a weight function by using the central point and radius. Moreover, they also gave the formula to calculate the central point and radius of fuzzy data. We will adopt the information to define fuzzy demand in our model. First, we introduce the following Lemma to calculate the central point and radius.

**Lemma 6.3** Let $A = [a, b, c]$ be a triangular fuzzy number, then the central point and radius are written as $o = \frac{a + b + c}{3}$ and $l = \frac{c - a}{4}$, respectively. (40)

**Proof**: The proof is given in the property 3.2 in chapter 3.

We explain some properties of probability distribution function in the following.

Let us discuss a continuous random variable $X$ that has gamma distribution with parameters $k > 0$ and $\theta > 0$, written as $X \sim \Gamma(k, \theta)$.

**Property 6.1** Let $U = nX, n > 0$ and $X \sim \Gamma(k, \theta)$, then $U \sim \Gamma(k, n\theta)$ and the expected value of $U$ is $nk\theta$. (1)

If $X \sim \Gamma(1, \lambda)$, we say that $X$ has an exponential distribution with mean $\lambda$. It is denoted as $X \sim \text{exp}(\lambda)$.

**Property 6.2** Let $U = nX, n > 0$ and $X \sim \text{exp}(\lambda)$, then $U \sim \text{exp}(n\lambda)$ and the expected value of $U$ is $n\lambda$. (1)

Assume that a continuous random variable $X$ has normal distribution with mean $\mu$ and variance $\sigma^2 > 0$, written as $X \sim N(\mu, \sigma^2)$. 

73
6. A PARAMETRIC ASSESSMENT APPROACH TO SOLVING FACILITY LOCATION PROBLEMS WITH FUZZY DEMANDS

Property 6.3 Let \( U = nX \), \( n > 0 \) and \( X \sim N(\mu, \sigma^2) \), then \( U \sim N(n\mu, n^2\sigma^2) \) and the expected value of \( U \) is \( n\mu \). (1)

Assume that a continuous random variable \( X \) has Weibull distribution with parameters \( w > 0 \) and \( \nu > 0 \), where \( w \) and \( \nu \) are the shape parameter and the scale parameter of the distribution, respectively. It is denoted as \( X \sim W(\nu, w) \).

Property 6.4 Let \( U = nX \), \( n > 0 \) and \( X \sim W(\nu, w) \), then \( U \sim W(n\nu, w) \) and the expected value of \( U \) is \( n\nu \Gamma(1 + \frac{1}{w}) \). (20)

In the next section, let us describe our facility location problem and explain the fuzzy demand in our model. Moreover, we will propose the procedure to solve the proposed model.

6.3 Problem Statements

Problem 1: Firm A intends to open facilities in \( n \) potential sites. The problem is to locate the facilities so as to maximize the total return from \( m \) clients. The cost of each facility consists of fixed opening cost and operating cost \( (c_i) \), variable operating cost \( (V_i) \) and transportation cost \( (t_{ij}) \) from \( i \) to \( j \). The \( m \) clients have total fuzzy demand \( D_m \) for the item provided by \( n \) facilities. Let us use the following notations.

Notations

\( i \): Index of facilities, \( 1 \leq i \leq n \).
\( j \): Index of clients, \( 1 \leq j \leq m \).
\( A_j \): The triangular fuzzy numbers \( A_j \) of client \( j \), where \( A_j = (a_j, b_j, c_j) \), \( \forall j = 1, 2, \ldots, m \).
\( o_j \): The central point of triangular fuzzy number \( A_j \).
\( l_j \): The radius of triangular fuzzy number \( A_j \).
\( E(A_j) \): The actual expected value of \( A_j \) (i.e. calculate the expected value from sample data).
\( EE(A_j) \): The estimated expected value (i.e. calculate the expected value from estimated parameters of underlying distribution function) of \( A_j \).
\( D_j \): Fuzzy demand of client \( j \).
\( RFD_j \): The realization of fuzzy demand \( D_j \) (i.e. defuzzify the fuzzy demand into a
6.3 Problem Statements

real number.)

$RFD_j^+$: The upper bound of $RFD_j$

$RFD_j^-$: The lower bound of $RFD_j$

$r_j$: Unit price charged to client $j$

$s_i$: Capacity of facility $i$

$c_i$: Fixed cost for opening and operating facility $i$

$V_i$: Unit variable operating cost of facility $i$

$x_i$: Binary decision variable equal to one if facility $i$ is open and zero otherwise.

$y_{ij}$: Quantity supplied to client $j$ from $i$

$t_{ij}$: Unit transportation cost from $i$ to $j$

$(1 - \theta_j)100\%$: The mean absolute percentage error of fuzzy demand (MAPE-FD) $D_j$

The facility location problem can be solved by setting the facility location model (6.1) as follows.

**Facility Location Model (6.1)**

\[
\begin{align*}
\text{max} \quad & \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} (r_j - V_i - t_{ij}) y_{ij} - \sum_{i=1}^{n} c_i x_i \right\} \\
\text{s.t.} \quad & \sum_{i=1}^{n} y_{ij} \leq D_j, \quad j = 1, 2, \cdots, m \\
& \sum_{j=1}^{m} y_{ij} \leq s_i x_i, \quad i = 1, 2, \cdots, n \\
& x_i \in \{0, 1\}, \quad i = 1, 2, \cdots, n; \\
& y_{ij} \geq 0, \quad j = 1, 2, \cdots, m.
\end{align*}
\]  

First, we give the definition of fuzzy demand in the following.

**Definition 6.2** Let $A = (b, L_1, L_2)$ be a triangular fuzzy number. The fuzzy demand of client $j$ is defined as follows.

\[
D_j = k N_j(EE(b_j), EE(L_{1j}), EE(L_{2j})),
\]

where $j = 1, 2, \cdots, m$, $k$ is the demand parameter, $N_j$ is the total number of population at cite $j$. When we know the probability distribution functions of $b_j$, $L_{1j}$ and $L_{2j}$, we can calculate the expected value through their parameters of underlying probability
distribution function by using Properties 6.1 to 6.4. Since the expected value is calculated by the estimated parameters of probability distribution function, we call it as the estimated expected value, denoted as "EE value".

To solve the facility location model (6.1), we defuzzify the fuzzy demand. Its defuzzification formula is defined by using central point and radius of triangular fuzzy data. Now, we give Theorem 6.1 to calculate the EE value of central point and radius.

**Theorem 6.1** Let \( A \equiv (b, L_1, L_2) \) be a triangular fuzzy number. The EE value of central point \( o \) and radius \( l \) are calculated by

\[
EE(o) = EE(b) + \frac{EE(L_2) - EE(L_1)}{3}
\]

and

\[
EE(l) = \frac{EE(L_1) + EE(L_2)}{4},
\]

respectively.

**Proof:** By Lemma 6.3, if \( A = [a, b, c] \) be a triangular fuzzy number, then we have the central point \( o = \frac{a + b + c}{3} \) and radius \( l = \frac{c - a}{4} \). Moreover, we can calculate \( L_1 = b - a \) and \( L_2 = c - b \).

Hence,

\[
EE(o) = EE(\frac{a + b + c}{3}) = EE(\frac{(b - L_1) + b + (b + L_2)}{3})
\]

\[
= EE(\frac{3b - L_1 + L_2}{3}) = EE(b) + \frac{EE(L_2) - EE(L_1)}{3},
\]

and

\[
EE(l) = EE(\frac{c - a}{4}) = EE(\frac{(b + L_2) - (b - L_1)}{4})
\]

\[
= EE(\frac{L_2 + L_1}{4}) = \frac{EE(L_1) + EE(L_2)}{4}.
\]

In this chapter, we define the fuzzy demand by using the estimated parameters of the probability distribution function of triangular fuzzy number. We test the parametric values of fuzzy demand by \( t \)-test which proposed in chapter 5. The results tell us that the fuzzy demand have some errors. To reduce the errors, we define a mean absolute percentage error of the fuzzy demand (MAPE-FD) in the following.

**Definition 6.3 Mean Absolute Percentage Error of the Fuzzy Demand (MAPE-FD)**

The mean absolute percentage error of the fuzzy demand (MAPE-FD) is a measure of accuracy of a method for constructing fitted values in statistics. It usually expresses
6.3 Problem Statements

accuracy as a percentage, and is defined by the formula:

\[
\text{MAPE - FD} = 100\% \frac{1}{n} \sum_{k=1}^{n} \left| \frac{E_k - EE_k}{E_k} \right|,
\]  

(6.2)

where \(E_k\) is the actual expected value and \(EE_k\) is the estimated expected value.

The absolute value in this calculation is summed up for every fitted or estimated point and divided again by the number \(n\) of fitted points. Multiplying by 100 makes a percentage error.

Using those concepts above, we define a method for defuzzifying the fuzzy demand and consider the MAPE-FD in this defuzification formula. We call it as realization of the fuzzy demand (RFD) defined as follows.

**Definition 6.4 Realization of the Fuzzy Demand (RFD)**

Let \(A \equiv (b, L_1, L_2)\) be a triangular fuzzy value on \(U\), which is the universal space. The realization of fuzzy demand is defined as follows:

\[
RFD_j = k N_j [\theta_j 100\% RFD_j^+ + (1 - \theta_j) 100\% RFD_j^-],
\]  

(6.3)

where

\[
RFD_j^+ = k N_j \{EE(o_j) + [1 - e^{-EE(l_j)}]\},
\]  

(6.4)

and

\[
RFD_j^- = k N_j \{EE(o_j) - [1 - e^{-EE(l_j)}]\},
\]  

(6.5)

here, \(EE(o_j)\) and \(EE(l_j)\) are the EE values of central point and radius, respectively, which can be calculated by Theorem 6.1, where \(k\) is the demand parameter, \(N_j\) is the total number of populations and \((1 - \theta_j) 100\%\) is the mean absolute percentage error of the fuzzy demand (MAPE-FD), \(0 \leq \theta_j \leq 1, \forall j = 1, 2, \ldots, m\).

Note that the realization of fuzzy demand \(RFD_j\) is composed of \(RFD_j^+\) and \(RFD_j^-\), and \(RFD_j^+\) and \(RFD_j^-\) is composed of \(EE(o_j)\) and \(1 - e^{-EE(l_j)}\). Hence, the MAPE-FD is calculated by the MAPE of \(EE(o_j)\) and MAPE of \(1 - e^{-EE(l_j)}\).

That is

\[
\text{MAPE - FD} = (1 - \theta_j) 100\% = \frac{100\%}{2} \left( | \frac{EE(o_j) - EE(o_j)}{E(o_j)} | + | \frac{1 - e^{-EE(l_j)}}{E(o_j)} - [1 - e^{-EE(l_j)}] | \right),
\]  

77
where \( E(o_j) \) and \( E(l_j) \) are the actual expected values which are calculated from sample data, and \( EE(o_j) \) and \( EE(l_j) \) are the EE value of central point and radius, respectively.

**Explanation of Definition 6.4**

To solve the facility location model (6.1), we first have to defuzzify the fuzzy demand. We define a function \( RFD_j \), which is calculated by \( RFD_j^+ \) and \( RFD_j^- \). The parameters in \( RFD_j^+ \) and \( RFD_j^- \) are calculated by the EE value of central point \( o_j \) and radius \( l_j \).

Note that \( EE(o_j) \) is the EE value of central point \( o_j \) in fuzzy demand of triangular fuzzy numbers. Moreover, we give another formula \( 1 - e^{-EE(l_j)} \), which is an increasing function. This function can provide a few quantities about the length from central point to left end point and right end point of a triangular fuzzy number. We define the upper bound demand and the lower bound demand from the EE value of central point \( o_i \) to add this quantity \( 1 - e^{-EE(l_j)} \) and the EE value of central point \( o_i \) to subtract this quantity \( 1 - e^{-EE(l_j)} \), respectively. Hence, the upper bound demand \( RFD_j^+ \) is defined by multiplying the quantity \( EE(o_j) + [1 - e^{-EE(l_j)}] \) and the lower bound demand \( RFD_j^- \) is defined by multiplying \( EE(o_j) - [1 - e^{-EE(l_j)}] \).

In Equation (6.3), we define the realization of fuzzy demand \( RFD_j \) which is composed of the upper bound demand \( RFD_j^+ \) and lower bound demand \( RFD_j^- \). Since the values of \( EE(o_j) \) and \( EE(l_j) \) will reduce a number of errors by using the estimated parameters. Hence, the mean absolute percentage error of fuzzy demand (MAPE-FD) \((1 - \theta_j)100\% \) is given here to calculate the actual demand we need.

The actual demand can be calculated by multiplying \( \theta_j100\% \), which indicates the percentage numbers within upper bound and lower bound demands. The calculation is as follows:

\[
kN_jRFD_j^- + \theta_j100\% \times kN_j[RFD_j^+ - RFD_j^-] \\
= kN_j[\theta_j100\%RFD_j^+ + (1 - \theta_j)100\%RFD_j^-].
\]

Hence, we got the formula (6.3).
Now, the facility location problem can be solved as a conventional optimal model. We rewrite facility location model (6.1) as follows.

**Facility Location Model (6.7)**

\[
\begin{align*}
\max \ & \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} \left[ (r_j - V_i - t_{ij}) y_{ij} \right] - \sum_{i=1}^{n} c_i x_i \right\} \\
\text{s.t.} \ & \sum_{i=1}^{n} y_{ij} \leq RFD_j, \quad j = 1, 2, \ldots, m \\
& \sum_{j=1}^{m} y_{ij} \leq s_i x_i, \quad i = 1, 2, \ldots, n \\
& x_i \in \{0, 1\}, \quad i = 1, 2, \ldots, n; \\
& y_{ij} \geq 0, \quad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, m.
\end{align*}
\]

(6.7)

Here, the procedure can be given in the following to solve a facility location problem with triangular fuzzy data.

**Step 1:** To create fuzzy questionnaires and ask clients to answer each question in real numbers.

**Step 2:** Get the triangular fuzzy numbers \((a_k^k, b_k^k, c_k^k)\), \(\forall k = 1, 2, \ldots, n\).

**Step 3:** Compute three parameters \(b_k^k, L_1^k, L_2^k\) from the triangular fuzzy numbers, \(\forall k = 1, 2, \ldots, n\).

**Step 4:** Calculate the probability distribution functions by simulating the parameters \(b_k^k, L_1^k, L_2^k\), \(\forall k = 1, 2, \ldots, n\), individually.

**Step 5:** Get the probability distribution functions of \(b_j, L_{1j}, L_{2j}\), \(\forall j = 1, 2, \cdots, m\).

**Step 6:** Calculate the EE values of \(b_j, L_{1j}, L_{2j}\), \(\forall j = 1, 2, \cdots, m\), by properties 6.1 to 6.4.

**Step 7:** Calculate the EE values of central point \(o_j\) and radius \(l_j\), \(\forall j = 1, 2, \cdots, m\), by Theorem 6.1.

**Step 8:** Calculate the MAPE – FD\_j by formula (6.6) and \(RFD_j\), \(\forall j = 1, 2, \cdots, m\), by Definition 4.
Step 9: Collect the other parameters we need in facility location model (6.7).

Step 10: Solve the facility location model (6.7) by using GP-IGP (Linear and Integer Goal Programming).

Now, we give an empirical study in the following section.

6.4 Empirical Study

Example 6.1 A frozen food Company makes a plan to set up 10 facilities in Taipei city, Taiwan. The company sells frozen fried chicken for Kentucky Fried Chicken (KFC). A manager assumes that when KFC is near the college, it will get the best return. Hence, they focus on surveying the demand of KFC near 5 colleges in Taipei city, Taiwan. The problem is to decide where the facilities should be set up to maximize the total return from 5 KFCs near 5 colleges in Taipei city, Taiwan. We chose 5 KFCs near the 5 colleges: Shih Chien University (Taipei), National Chungchi University, Soochow University, Chinese Culture University, and National Taipei University of Education, and denoted these 5 KFCs as $K_1, K_2, K_3, K_4$ and $K_5$, respectively.

The KFC’s primary product is pressure-fried pieces of chicken made from an "original recipe". The company also sells burgers, wraps and a variety of finger foods, including chicken strips, wings, nuggets, and popcorn chicken. We focused on the fried chicken meals in our facility location problem. The fried chickens were supplied by 10 facilities (frozen food plants) whose 10 potential sites are plotted in Figure 6.1. Our task is to locate the optimal sites for the fried chicken supplier (frozen food plants) from all the 10 potential alternatives.

To know the fuzzy demand in those 5 KFCs, we started to survey the demand of client $j$. We chose 100 peoples who entered the 5 KFCs and asked them the following questions as follows.

1. How many times (interval) do you buy the fried chicken meals at KFCs per week?
2. How many times (real numbers) do you most possibly buy the fried chicken meals at KFCs per week?

We gathered those data and got the triangular fuzzy numbers. Table 6.1 shows the results.
Figure 6.1: The Potential Facility Sites and Universities in Taipei City

Table 6.1: Triangular Fuzzy Numbers \((a^k, b^k, c^k)\) from Five KFCs.

<table>
<thead>
<tr>
<th></th>
<th>((a^k, b^k, c^k))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_1)</td>
<td>(2,3,6) (3,3,4) (\cdots) (1,2,2)</td>
</tr>
<tr>
<td>(K_2)</td>
<td>(2,2,5) (1,1,2) (\cdots) (3,5,7)</td>
</tr>
<tr>
<td>(K_3)</td>
<td>(6,6,7) (3,3,5) (\cdots) (2,3,5)</td>
</tr>
<tr>
<td>(K_4)</td>
<td>(1,3,5) (1,3,4) (\cdots) (1,2,3)</td>
</tr>
<tr>
<td>(K_5)</td>
<td>(1,1,1) (1,1,3) (\cdots) (2,2,5)</td>
</tr>
</tbody>
</table>
6. A PARAMETRIC ASSESSMENT APPROACH TO SOLVING FACILITY LOCATION PROBLEMS WITH FUZZY DEMANDS

After we obtained data, we can calculate the middle point $b^k$, left width $L^k_1$ and right width $L^k_2$ by Definition 6.1. Table 6.2 shows the results.

Table 6.2: Three Parameters $(b^k, L^k_1, L^k_2)$ of Triangular Fuzzy Numbers from 5 KFCs.

<table>
<thead>
<tr>
<th>K</th>
<th>$(b^k, L^k_1, L^k_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_1$</td>
<td>(2,1,3) (3,0,1) ··· (2,1,0)</td>
</tr>
<tr>
<td>$K_2$</td>
<td>(2,0,3) (1,0,1) ··· (5,2,2)</td>
</tr>
<tr>
<td>$K_3$</td>
<td>(6,0,1) (3,0,2) ··· (3,1,2)</td>
</tr>
<tr>
<td>$K_4$</td>
<td>(3,2,2) (3,2,1) ··· (2,1,1)</td>
</tr>
<tr>
<td>$K_5$</td>
<td>(1,0,0) (1,0,2) ··· (2,0,3)</td>
</tr>
</tbody>
</table>

Now, we found the probability distribution functions by simulating $b^k$, $L^k_1$ and $L^k_2$, individually. Table 6.3 shows the results.

Table 6.3: The Probability Distribution Functions for Parameters $b_j$, $L_{1j}$ and $L_{2j}$.

<table>
<thead>
<tr>
<th>$b_j$</th>
<th>$L_{1j}$</th>
<th>$L_{2j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_1$</td>
<td>$N(2.45, 0.63^2)$</td>
<td>$W(2.50, 0.89^2)$</td>
</tr>
<tr>
<td>$K_2$</td>
<td>$W(2.34, 2.99^2)$</td>
<td>$W(1.84, 1.07^2)$</td>
</tr>
<tr>
<td>$K_3$</td>
<td>$\Gamma (4.70, 0.56^2)$</td>
<td>$N(0.86, 0.41^2)$</td>
</tr>
<tr>
<td>$K_4$</td>
<td>$W(2.94, 2.63^2)$</td>
<td>$N(1.03, 0.49^2)$</td>
</tr>
<tr>
<td>$K_5$</td>
<td>$\Gamma (3.13, 0.68^2)$</td>
<td>$\Gamma (1.94, 0.26^2)$</td>
</tr>
</tbody>
</table>

When we know those probability distribution functions from three parameters of triangular fuzzy numbers, we can calculate the EE value based on the parameters of probability distribution functions by using Properties 6.1 to 6.4. Table 6.4 shows the results.

Since each fried chicken meal served by the KFCs has 80g meat in average, and one year consists of 52 weeks, the fuzzy demand (for fried chicken meals per ton) of each student for the whole year can be written as follows.

$$D_j = 4.16 \times 10^{-3} N_j(EE(b_j), EE(L_{1j}), EE(L_{2j})),$$

(6.8)

for $j = 1, 2, \cdots, 5$. Moreover, we calculated the EE values of central point $o_j$ and radius $l_j, \forall j = 1, 2, \cdots, m$, by Theorem 6.1.
6.4 Empirical Study

Table 6.4: The EE Value of \( b_j \), \( L_{1j} \), and \( L_{2j} \).

<table>
<thead>
<tr>
<th></th>
<th>( EE(b_j) )</th>
<th>( EE(L_{1j}) )</th>
<th>( EE(L_{2j}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_1 )</td>
<td>2.45</td>
<td>2.65</td>
<td>1.28</td>
</tr>
<tr>
<td>( K_2 )</td>
<td>2.09</td>
<td>1.79</td>
<td>1.63</td>
</tr>
<tr>
<td>( K_3 )</td>
<td>2.63</td>
<td>0.86</td>
<td>1.98</td>
</tr>
<tr>
<td>( K_4 )</td>
<td>2.61</td>
<td>1.03</td>
<td>1.02</td>
</tr>
<tr>
<td>( K_5 )</td>
<td>2.13</td>
<td>0.50</td>
<td>2.18</td>
</tr>
</tbody>
</table>

To ensure the accuracy of EE values of central point \( o_j \) and radius \( l_j \), we use \( t \)-test, which is built in chapter 5, to test the accuracy of EE values first. We give the results in Table 6.5.

Table 6.5: Decision by \( t \)-test

<table>
<thead>
<tr>
<th></th>
<th>( K_1 )</th>
<th>( K_2 )</th>
<th>( K_3 )</th>
<th>( K_4 )</th>
<th>( K_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( EE )</td>
<td>(1.99, 0.98)</td>
<td>(2.04, 0.86)</td>
<td>(3.00, 0.71)</td>
<td>(2.61, 0.51)</td>
<td>(2.69, 0.67)</td>
</tr>
<tr>
<td>( E )</td>
<td>(2.62, 0.49)</td>
<td>(2.92, 0.59)</td>
<td>(2.78, 0.48)</td>
<td>(2.30, 0.44)</td>
<td>(2.67, 0.60)</td>
</tr>
<tr>
<td>( S_n )</td>
<td>(0.68, 0.20)</td>
<td>(1.19, 0.29)</td>
<td>(1.30, 0.22)</td>
<td>(0.97, 0.19)</td>
<td>(1.48, 0.41)</td>
</tr>
<tr>
<td>( A )</td>
<td>[2.49, 2.76]</td>
<td>[2.68, 3.15]</td>
<td>[2.52,3.04]</td>
<td>[2.11, 2.49]</td>
<td>[2.38,2.96]</td>
</tr>
<tr>
<td></td>
<td>( \times[0.46,0.53] )</td>
<td>( \times[0.53,0.65] )</td>
<td>( \times[0.44,0.52] )</td>
<td>( \times[0.40,0.48] )</td>
<td>( \times[0.51,0.68] )</td>
</tr>
<tr>
<td>Decision</td>
<td>( EE \notin A )</td>
<td>( EE \notin A )</td>
<td>( EE \notin A )</td>
<td>( EE \notin A )</td>
<td>( EE \in A )</td>
</tr>
</tbody>
</table>

We also give the percentage error of central point \( o_i \) and radius \( l_i \) in Table 6.6.

Table 6.6: Percentage Error of \( EE(o_j) \) and \( EE(l_j) \)

<table>
<thead>
<tr>
<th></th>
<th>( E(o_j) - EE(o_j) )</th>
<th>Decision of ( EE(o_j) )</th>
<th>( E(l_j) - EE(l_j) )</th>
<th>Decision of ( EE(l_j) )</th>
<th>Decision of ( EE )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_1 )</td>
<td>24.05%</td>
<td>( EE(o_1) \notin A )</td>
<td>100.00%</td>
<td>( EE(l_1) \notin A )</td>
<td>( EE \notin A )</td>
</tr>
<tr>
<td>( K_2 )</td>
<td>30.14%</td>
<td>( EE(o_2) \notin A )</td>
<td>45.76%</td>
<td>( EE(l_2) \notin A )</td>
<td>( EE \notin A )</td>
</tr>
<tr>
<td>( K_3 )</td>
<td>7.91%</td>
<td>( EE(o_3) \in A )</td>
<td>47.92%</td>
<td>( EE(l_3) \notin A )</td>
<td>( EE \notin A )</td>
</tr>
<tr>
<td>( K_4 )</td>
<td>13.48%</td>
<td>( EE(o_4) \notin A )</td>
<td>15.91%</td>
<td>( EE(l_4) \notin A )</td>
<td>( EE \notin A )</td>
</tr>
<tr>
<td>( K_5 )</td>
<td>0.74%</td>
<td>( EE(o_5) \in A )</td>
<td>11.66%</td>
<td>( EE(l_5) \notin A )</td>
<td>( EE \in A )</td>
</tr>
</tbody>
</table>

We can see that when the percentage error of \( EE(o_j) \) and percentage error of \( EE(l_j) \) are small, we do not reject the EE value. On the other hand, if we reject the EE value, it means that we have large percentage error. Hence, if we want to defuzzify fuzzy
6. A PARAMETRIC ASSESSMENT APPROACH TO SOLVING FACILITY LOCATION PROBLEMS WITH FUZZY DEMANDS

demand by using those EE values, we must consider the concept of MAPE-FD in the defuzzification formula.

Hence, we calculate all parameter values we need in Definition 6.4. Table 6.7 shows the parameter values.

<table>
<thead>
<tr>
<th>$N_j$</th>
<th>$EE(o_j)$</th>
<th>$EE(l_j)$</th>
<th>$1 - e^{-EE(l_j)}$</th>
<th>$RFD^-_j$</th>
<th>$RFD^+_j$</th>
<th>$RFD_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_1$</td>
<td>15,600</td>
<td>1.99</td>
<td>0.98</td>
<td>0.63</td>
<td>170</td>
<td>89</td>
</tr>
<tr>
<td>$K_2$</td>
<td>15,800</td>
<td>2.04</td>
<td>0.86</td>
<td>0.57</td>
<td>172</td>
<td>96</td>
</tr>
<tr>
<td>$K_3$</td>
<td>15,800</td>
<td>3.00</td>
<td>0.71</td>
<td>0.51</td>
<td>231</td>
<td>164</td>
</tr>
<tr>
<td>$K_4$</td>
<td>26,900</td>
<td>2.61</td>
<td>0.51</td>
<td>0.40</td>
<td>337</td>
<td>247</td>
</tr>
<tr>
<td>$K_5$</td>
<td>5,900</td>
<td>2.69</td>
<td>0.67</td>
<td>0.49</td>
<td>78</td>
<td>54</td>
</tr>
</tbody>
</table>

For the other parameter values in the model (6.7), since all the sites as well as KFCs are located within the area of Taipei (including Taipei City and Taipei County), the average transportation cost of one ton frozen foods from each site to each KFC is around $t_{ij} = 1,600$ NTD, equally for $i = 1, 2, \cdots, 10$, $j = 1, 2, \cdots, 5$. Finally, the unit price $r_j$ of one ton meat charged to client $j$ (KFC) is 150,000 NTD, equally for $j = 1, 2, \cdots, 5$. The other parameter values are shown in Table 6.8.

With the obtained realization of fuzzy demand $RFD_j$ in Definition 6.4, $\forall j = 1, 2, \cdots, 5$, the average of interval values for unit variable cost $V_i$ of each plant, and the other parameter values mentioned in Table 6.8, the frozen food plant location problem
Table 6.8: The Fixed Cost, Variable Cost and Capacity for Each Potential Site.

<table>
<thead>
<tr>
<th>Site</th>
<th>Capacity ($s_i$ ton)</th>
<th>Fixed cost ($c_i$ $10^6$ NTD)</th>
<th>Unit variable cost ($V_i$ $10^3$ NTD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4,000</td>
<td>43.85</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>1,500</td>
<td>105.62</td>
<td>110</td>
</tr>
<tr>
<td>3</td>
<td>1,800</td>
<td>53.12</td>
<td>110</td>
</tr>
<tr>
<td>4</td>
<td>1,300</td>
<td>101.80</td>
<td>110</td>
</tr>
<tr>
<td>5</td>
<td>1,000</td>
<td>29.30</td>
<td>110</td>
</tr>
<tr>
<td>6</td>
<td>3,400</td>
<td>113.72</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>3,700</td>
<td>44.55</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>1,800</td>
<td>44.12</td>
<td>110</td>
</tr>
<tr>
<td>9</td>
<td>4,000</td>
<td>41.85</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>4,200</td>
<td>51.60</td>
<td>100</td>
</tr>
</tbody>
</table>

can be formulated by using facility location model (6.7) as follows:

\[
\begin{align*}
\max & \quad \left\{ \sum_{i=1}^{10} \sum_{j=1}^{5} [(r_{ij} - V_i - t_{ij})y_{ij}] - \sum_{i=1}^{10} c_i x_i \right\} \\
\text{s.t.} & \quad \sum_{i=1}^{10} y_{ij} \leq RFD_j, \quad j = 1, 2, \ldots, 5 \\
& \quad \sum_{j=1}^{5} y_{ij} \leq s_i x_i, \quad i = 1, 2, \ldots, 10 \\
& \quad x_i \in \{0, 1\}, \quad i = 1, 2, \ldots, 10; \\
& \quad y_{ij} \geq 0, \quad i = 1, 2, \ldots, 10; \quad j = 1, 2, \ldots, 5.
\end{align*}
\]

(6.9)

We solved the model (6.9) by using GP-IGP (Linear and Integer Goal Programming) and obtained the optimal solution as follows:

\[x^* = (0, 0, 0, 0, 1, 0, 0, 0, 0, 0),\]

with the objective value (Max. Profit) 5.759 million NTD. This means the site 5 in TuCheng City is an optimal potential site for the new frozen food plant.
6. A PARAMETRIC ASSESSMENT APPROACH TO SOLVING
FACILITY LOCATION PROBLEMS WITH FUZZY DEMANDS

6.5 Summary of Chapter 6

6.5.1 Comparison

Many research studies (see the discussion in (12), (51), (52) and (67)) have adopted
the fuzzy demand in terms of location problems. However, these studies have all used
only the central point to defuzzify the fuzzy demand. In Cheng’s (8) paper, the dis-
tance method was used to defuzzify fuzzy data. The distance formula includes two
parameters, an \( \bar{x}_i \) value on the horizontal axis and a \( \bar{y}_i \) value on the vertical axis, which
is very similar to our method of using two parameters to address the problem. (Note
that we used the EE value of central point and the EE value of radius to define the
defuzzification formula.) When we used the distance method to defuzzify the fuzzy de-
mmand in our location problems and took the mean value of \( V_i \), we obtained a maximum
profit (here called the distance solution (DS)) about 4.530 million NTD.

Note that we used the membership function \( f(x) \) as specified in Definition 6.1. The
fuzzy demand

\[
D_j = 4.16 \times 10^{-3} N_j (EE(b_j), EE(L_{1j}), EE(L_{2j})),
\]

implied

\[
D_j' = 4.16 \times 10^{-3} N_j \sqrt{E(\bar{x}_i)^2 + E(\bar{y}_i)^2},
\]

where \( \bar{x}_i = o_i \) and \( \bar{y}_i = \frac{\int_0^1 y g_X^L dy + \int_1^0 y g_X^R dy}{\int_0^1 g_X^L dy + \int_1^0 g_X^R dy} = \frac{a + 4b + c}{3(a + 2b + c)} \), where \( g_X^L = a + (b - a)y \)
and \( g_X^R = c + (b - c)y \) are inverse functions \( \forall y \in [0, 1] \).

We computed the parameter values in Table 6.9, which were used in the distance
method.

<table>
<thead>
<tr>
<th>( N_j )</th>
<th>( E(\bar{x}_i) )</th>
<th>( E(\bar{y}_i) )</th>
<th>( \sqrt{E(\bar{x}_i)^2 + E(\bar{y}_i)^2} )</th>
<th>( D_j' )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_1 )</td>
<td>15,600</td>
<td>2.62</td>
<td>0.49</td>
<td>2.67</td>
</tr>
<tr>
<td>( K_2 )</td>
<td>15,800</td>
<td>2.92</td>
<td>0.49</td>
<td>2.96</td>
</tr>
<tr>
<td>( K_3 )</td>
<td>15,800</td>
<td>2.78</td>
<td>0.49</td>
<td>2.82</td>
</tr>
<tr>
<td>( K_4 )</td>
<td>26,900</td>
<td>2.30</td>
<td>0.51</td>
<td>2.34</td>
</tr>
<tr>
<td>( K_5 )</td>
<td>5,900</td>
<td>2.67</td>
<td>0.47</td>
<td>2.71</td>
</tr>
</tbody>
</table>
6.5 Summary of Chapter 6

In this chapter, we have defined a fuzzy demand in our facility location model. We defuzzified the fuzzy demand by calculating the value of $RFD_j$ in formula (6.3) in Definition 6.4. The RFD formula combines the $RFD^+$ and $RFD^-$. The most important parts of composition of $RFD^+$ and $RFD^-$ are the EE value of the central point and the EE value of the radius. Moreover, the MAPE-FD formula has been proposed in the RFD formula. We took the mean value of the fuzzy unit variable costs $V_i$. We obtained a maximum profit (here called the realization of the fuzzy demand solution (RFDS)) about 5.759 million NTD.

When we defuzzified the fuzzy demand by calculating the $RFD_j^+$ in formula (6.4) and took the mean value of $V_i$, we obtained a maximum profit (here called the upper bound of the RFD solution (RFDS+)) about 8.639 million NTD. Moreover, if we defuzzify the fuzzy demand by calculating the $RFD_j^-$ in formula (6.5) and take the mean value of $V_i$. We obtained a maximum profit (here called the lower bound of the RFD solution (RFDS-)) about -4.340 million NTD. It means that we should not build the facility because the demand is not enough to pay the facility’s fixed costs.

When we used the EE value by choosing the middle point of the triangular fuzzy numbers as the parameter of the demand of client $j$, took the number $4.16 \times 10^{-3}N_j EE(b_j)$ instead of $D_j$ in facility location model (6.1) and took the mean value of $V_i$, we obtained a maximum profit (here called the mean demand solution (MDS)) about 1.919 million NTD.

Moreover, when we took the EE value by choosing the central point of the triangular fuzzy numbers as the parameter of the demand of client $j$, took the number $4.16 \times 10^{-3}N_j EE(o_j)$ instead of $D_j$ in the facility location model (6.1) and took the mean value of $V_i$, we obtained a maximum profit (here called the central demand solution (CDS)) about 2.111 million NTD. We listed all of the solutions in Table 6.10 for comparison.

From Table 6.10, we see that the optimal solution indicated by all of the approaches is Site 5 in TuCheng City. We calculated the EE value of the $RFDS^+$, RFDS, $RFDS^-$, MDS and CDS methods. We calculated the expected value directly from sample data of the DS method. Comparing these approach methods, the conventional defuzzification methods (the MDS and CDS methods) defuzzified the fuzzy data into a single point. Hence, the maximum profit under these methods were smaller than those in the $RFDS^+$, RFDS and DS methods. Therefore, we propose consideration of the other
6. A PARAMETRIC ASSESSMENT APPROACH TO SOLVING FACILITY LOCATION PROBLEMS WITH FUZZY DEMANDS

Table 6.10: Solution Comparisons of Location Problem

<table>
<thead>
<tr>
<th>Approach Method</th>
<th>Optimal Solution</th>
<th>( D_j )</th>
<th>( V_i )</th>
<th>Max. Profit (10^6 NTD)</th>
<th>Percentage Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS (0,0,0,1,0,0,0,0)</td>
<td>( 4.16 \times 10^{-3} N_j \sqrt{E(\bar{x}_i)^2 + E(\bar{y}_i)^2} )</td>
<td>( \frac{V_i^c + V_i^l}{2} )</td>
<td>4.530</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( RFDS^+ ) (0,0,0,0,1,0,0,0,0)</td>
<td>( RFD_{ij}^+ )</td>
<td>( \frac{V_i^c}{2} )</td>
<td>8.639</td>
<td>90.70%</td>
<td></td>
</tr>
<tr>
<td>RFDS (0,0,0,0,0,1,0,0,0,0)</td>
<td>( RFD_j )</td>
<td>( \frac{V_i^c + V_i^l}{2} )</td>
<td>5.759</td>
<td>27.13%</td>
<td></td>
</tr>
<tr>
<td>( RFDS^- ) (0,0,0,0,0,0,1,0,0,0,0)</td>
<td>( RFD_{ij}^- )</td>
<td>( \frac{V_i^c}{2} )</td>
<td>-4.340</td>
<td>195.80%</td>
<td></td>
</tr>
<tr>
<td>MDS (0,0,0,0,0,1,0,0,0,0)</td>
<td>( 4.16 \times 10^{-3} N_j EE(b_j) )</td>
<td>( \frac{V_i^c + V_i^l}{2} )</td>
<td>1.919</td>
<td>57.63%</td>
<td></td>
</tr>
<tr>
<td>CDS (0,0,0,0,0,0,1,0,0,0,0)</td>
<td>( 4.16 \times 10^{-3} N_j EE(o_j) )</td>
<td>( \frac{V_i^c + V_i^l}{2} )</td>
<td>2.111</td>
<td>53.40%</td>
<td></td>
</tr>
</tbody>
</table>

parameters \((r_i)\) in the defuzzification formula in this chapter. The \( RFDS^+ \), RFDS, \( RFDS^- \) and DS methods all considered the other parameters \((\bar{y}_i \text{ and } r_i)\). The difference between the \( RFDS^+ \), RFDS, and \( RFDS^- \) methods is that we calculated the EE value in the \( RFDS^+ \), RFDS, and \( RFDS^- \) methods and we calculated the expected value directly from the sample data in the DS method. Although the maximum profit according to the \( RFDS^+ \) method is larger than that of the other methods, it requires too much demand in the location problem. Additionally, if we assess too little demand, the production will be insufficient, and we will obtain a negative profit based on the \( RFDS^- \) method. Moreover, the error percentage of our method (RFDS) based on the DS method is the smallest one. Hence, it is better to use the RFDS method to solve the facility location problem, considering the MAPE-FD formula in the RFD formula. The maximum profit in the RFDS method is close to that in the DS method. Thus, we have built the new frozen food plant at site 5, TuCheng City. The result of RFDS method surely achieved a larger profit than the result based on DS method.

6.5.2 Conclusions

It is often difficult to choose the most effective and efficient locations of individual facilities. Moreover, it is also difficult when using a critical procedure in the whole data-processing stage to determine the parameters for fuzzy random demands because the distribution function serves as a basis for computation.
Thus, we first identified the probability distribution function of our fuzzy data and calculated the EE value for the RFD formula. We combined the formulas $RFD_j^+$ and $RFD_j^-$ with the $MAPE - FD_j$ formula in the RFD formula. The value of $RFD_j$ was closer to the actual demand in each KFC.

In the entire process described above, the most important step was first finding the probability distribution function. If we did not know the probability distribution function, we could not determine the fuzzy demand in our facility location model. In this chapter, we dealt with a real-life location problem and used the parameter values of the probability distribution functions to calculate the EE values. The parameter values were provided in (39) and (72). We solved the facility location model (6.1) by defuzzifying the fuzzy demand. To defuzzify the fuzzy demand, we defined a defuzzification formula that includes an error assessment, the MAPE-FD formula, in the FRD formula. We also described an empirical study and compared various approach methods with our proposed method (i.e., the RFDS method). The results showed that we can obtain higher profits based on the approach of the RFDS method.

We would like to note some further points of improvement for the future.

1. In this chapter, we defined a fuzzy demand using triangular fuzzy numbers in our model. We estimated the expected value and defuzzified the fuzzy demand by determining the underlying probability distribution function of triangular fuzzy numbers and the RFD formula. Moreover, we included an error assessment with the MAPE-FD formula in the RFD formula. We hope that we can define a fuzzy demand with other types of fuzzy data and defuzzify them by improving the RFD formula in the future.

2. In this chapter, we set up many parameter values using the same values from facility $i$ to client $j$ in our model; for example, the average transportation cost, $t_{ij}$, of one ton of frozen food from each site to each KFC and the unit price, $r_j$, of one ton of meat charged to client $j$ (KFC). If we can obtain real data from real-life location problems in the future, it will make our results more realistic.
7

Conclusions and Future Work

7.1 Conclusions

In this thesis, we established a statistical test of fuzzy data that is called a fuzzy statistical test. We introduced a concept for "defuzzifying" fuzzy data into real numbers; that is, we used the central point and the radius instead of the fuzzy data itself. Because the central point and radius have the statistical characteristics of the mean and the variance, the conventional statistical tests can be applied.

In chapter 3, we studied the use of the K-S two-sample test with small samples of continuous fuzzy data. To identify the statistical pivot, we defined a new function, the weight function, which includes both the central point and the radius. The weight function could be used to classify all continuous fuzzy data. Moreover, we could divide fuzzy data samples into different classes. With this rule, the cumulative distribution function could be determined. As a result, we could obtain the statistical pivot of the K-S two-sample test with the continuous fuzzy data. We also provided empirical studies to show that fuzzy hypothesis testing with soft computing is a realistic and reasonable approach to dealing with continuous fuzzy data in the social science research.

In chapter 4, we continued to study the K-S two-sample test. We emphasized the distribution differences between two populations of fuzzy data. We presented a new function, $RF_x$, in this chapter, which consists of both the central points and the radius values. The $RF_x$ function could be used to classify all continuous fuzzy data. Moreover, we could differentiate two fuzzy samples. We also provided several empirical studies to compare the proposed method with conventional methods. Through this procedure,
an intelligent calculation method can be applied to analyze industrial, physiological, economic or financial data in the future.

To illustrate the efficacy of the proposed method, we introduced two real-life applications: portfolio selection problems in chapter 5 and facility location problems in chapter 6. In chapter 5, we introduced an application of maximizing an investment portfolio of foreign exchange currencies. The portfolio selection model was built by using the expected values and variances of the central point and radius. The expected values and variances were calculated using the estimated parameters of the underlying distribution function. We evaluated the best return using a fuzzy statistical test. In this procedure, from data extraction to the fuzzy statistical test, there is no doubt that the model can address interval data because we had ”defuzzified” the fuzzy data into real numbers before we solved the portfolio selection model. Hence, the model is comparable to a traditional linear programming model.

We provided the risk level $k$ for investors to make decisions. We needed to determine the value $k$ first and solve the linear programming model many times until we obtained the solution with only one exchange currency. Because we set the value of $k$ in the model, many expected returns were obtained that depend on the value of $k$. We obtained a maximum return with different risk levels in our model and made a decision for selecting the best return using a fuzzy statistical test. We concluded that it is more of a conservative investment and more objective for investors to make decision when they buy many exchange currencies. We also concluded that evaluation using the fuzzy statistical test enables us to obtain a stable expected return and a low-risk investment with different choices based on the risk level $k$.

In chapter 6, we described another real-life application to facility location problems. It is often difficult to choose the most effective and efficient locations of individual facilities. Moreover, it is also difficult for a critical procedure in the whole data-processing stage to determine the parameter for fuzzy demands because the distribution function serves as a basis for computation. From this view of our method, we first determined the probability distribution function of the fuzzy data and calculated the EE value for the RFD formula. We combined the formulas $RFD_j^+$ and $RFD_j^-$ with the $MAPE - FD_j$ formula in the RFD formula. The value of $RFD_j$ was closer to the actual demand at each KFC.
7. CONCLUSIONS AND FUTURE WORK

In the entire process above, the most important issue is first finding the probability distribution function. If we do not know the probability distribution function, we cannot determine the fuzzy demand in our facility location model. In this chapter, we dealt with a real-life location problem and used the parameter values of probability distribution functions to calculate the EE values. The parameter values were provided in the references (39) and (72). We solved the facility location model (6.1) by defuzzifying the fuzzy demand.

To defuzzify the fuzzy demand, we defined a defuzzification formula that includes an error assessment, MAPE-FD, in the FRD formula. We also conducted an empirical study and compared various other methods with our proposed method (i.e., the RFDS method). The results showed that we can obtain a higher profit using the RFDS method.

7.2 Future Work

Although studies in decision-making and evaluations with fuzzy statistical tests have been discussed in this thesis, there are still numerous further points in need of improvement by future research beyond the scope of this dissertation.

From a theoretical point of view,

1. We introduced a new concept for defuzzifying fuzzy data by using the central point and radius. We considered three types of fuzzy data individually: interval values, triangular fuzzy data, and trapezoidal fuzzy data. We propose that the proposed method would be more likely to be realized if we could analyze other types of fuzzy numbers with fuzzy statistical tests in the future.

2. The results based on a fuzzy statistical test can indicate two pieces of information, (for the example in chapter 5), as follows:

   (a) We tested the expected return using a fuzzy statistical test in chapter 5, and the results indicated whether we should accept or reject the expected return.

   (b) Because the expected return was solved based on the portfolio selection model and the parameters in the model were calculated using the estimated parameters of the underlying distribution function, we concluded that if we
accept the hypothesis (i.e., the expected return), there are no problems with
using data extraction to obtain an expected return based on the value of \( k \).

From the perspective of real-life application,

1. In the proposed portfolio selection model in chapter 5, we presented a constraint
inequality with a risk level \( k \) that was given in advance. We made the value of \( k \)
greater than zero and accurate to the second decimal place. We thought that we
could give more accurate risk levels and results for investors to make decisions in
the future.

2. We defined a fuzzy demand with triangular fuzzy numbers in our model in chapter
6. We estimated the expected value and defuzzified the fuzzy demand by find-
ing out the underlying probability distribution function of the triangular fuzzy
numbers and the RFD formula. Moreover, we included an error assessment with
the MAPE-FD formula in the RFD formula. We hope that we can define a fuzzy
demand with other types of fuzzy data and defuzzify it by improving the RFD
formula.

3. The financial market is in fact affected by many non-probabilistic factors, and
future returns of risky assets cannot be predicted accurately in any uncertain
economic environment. Although we can evaluate and select the best return
using a fuzzy statistical test, we also need to consider financial reports, experts’
individual experiences and other factors in real world. Moreover, if we can get
real data from real-life location problems in the future, our results will be more
realistic for decision-making processes.
Bibliography


BIBLIOGRAPHY


## Appendix A

Table I Kolmogorov-Smirnov Two-Sample Statistic (63)

<table>
<thead>
<tr>
<th>Number of Samples $m = n$</th>
<th>Significant Level $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.200</td>
</tr>
<tr>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>66</td>
</tr>
<tr>
<td>12</td>
<td>72</td>
</tr>
<tr>
<td>13</td>
<td>75</td>
</tr>
<tr>
<td>14</td>
<td>84</td>
</tr>
<tr>
<td>15</td>
<td>90</td>
</tr>
<tr>
<td>16</td>
<td>112</td>
</tr>
<tr>
<td>17</td>
<td>119</td>
</tr>
<tr>
<td>18</td>
<td>126</td>
</tr>
<tr>
<td>19</td>
<td>133</td>
</tr>
<tr>
<td>20</td>
<td>140</td>
</tr>
</tbody>
</table>

* The table gives the critical value of $mnD_{m,n}$ and its significance level $\alpha$ is given on the top row.

For $m$ and $n$ are large, the approximate critical value can be found by following formula.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>.200</th>
<th>.100</th>
<th>.050</th>
<th>.020</th>
<th>.010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$1.07\sqrt{\frac{m+n}{mn}}$</td>
<td>$1.22\sqrt{\frac{m+n}{mn}}$</td>
<td>$1.36\sqrt{\frac{m+n}{mn}}$</td>
<td>$1.52\sqrt{\frac{m+n}{mn}}$</td>
<td>$1.63\sqrt{\frac{m+n}{mn}}$</td>
</tr>
</tbody>
</table>

100
List of Publications

Peer-reviewed International Journals


Peer-reviewed International Conference Papers


**Book Chapter**
