Hung T. Nguyen, Berlin Wu

Fundamentals of Statistics with Fuzzy Data

Studies in Fuzziness and Soft Computing, Volume 198

Editor-in-chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series
can be found on our homepage:
springer.com

Vol. 182. John N. Mordeson,
Kiran R. Bhutani, Azriel Rosenfeld
Fuzzy Group Theory, 2005
ISBN 3-540-25072-7

Vol. 183. Larry Bull, Tim Kovacs (Eds.)
Foundations of Learning Classifier Systems, 2005
ISBN 3-540-25073-5

Vol. 184. Barry G. Silverman, Ashlesha Jain,
Ajita Ichalkaranje, Lakhmi C. Jain (Eds.)
Intelligent Paradigms for Healthcare Enterprises, 2005
ISBN 3-540-22903-5

Vol. 185. Spiros Sirmakessis (Ed.)
Knowledge Mining, 2005
ISBN 3-540-25070-0

Vol. 186. Radim Bělohlávek, Vilém Vychodil
Fuzzy Equational Logic, 2005
ISBN 3-540-26254-7

Vol. 187. Zhong Li, Wolfgang A. Halang,
Guanrong Chen (Eds.)
Integration of Fuzzy Logic and Chaos Theory, 2006
ISBN 3-540-26899-5

Vol. 188. James J. Buckley, Leonard J. Jowers
Simulating Continuous Fuzzy Systems, 2006
ISBN 3-540-28455-9

Vol. 189. Hans Bandemer
Mathematics of Uncertainty, 2006
ISBN 3-540-28457-5

Vol. 190. Ying-ping Chen
Extending the Scalability of Linkage Learning Genetic Algorithms, 2006
ISBN 3-540-28459-1

Vol. 191. Martin V. Butz
Rule-Based Evolutionary Online Learning Systems, 2006
ISBN 3-540-25379-3

Vol. 192. Jose A. Lozano, Pedro Larrañaga,
Iñaki Inza, Endika Bengoetxea (Eds.)
Towards a New Evolutionary Computation, 2006
ISBN 3-540-29006-0

Vol. 193. Ingo Glöckner
Fuzzy Quantifiers: A Computational Theory, 2006
ISBN 3-540-29634-4

Vol. 194. Dawn E. Holmes, Lakhmi C. Jain (Eds.)
Innovations in Machine Learning, 2006
ISBN 3-540-30609-9

Vol. 195. Zongmin Ma
Fuzzy Database Modeling of Imprecise and Uncertain Engineering Information, 2006
ISBN 3-540-30675-7

Vol. 196. James J. Buckley
Fuzzy Probability and Statistics, 2006
ISBN 3-540-30841-5

Vol. 197. Enrique Herrera-Viedma, Gabriella
Pasi, Fabio Crestani (Eds.)
Soft Computing in Web Information Retrieval, 2006
ISBN 3-540-31588-8

Vol. 198. Hung T. Nguyen, Berlin Wu
Fundamentals of Statistics with Fuzzy Data, 2006
ISBN 3-540-31695-7

Hung T. Nguyen
Berlin Wu

Fundamentals of Statistics with Fuzzy Data

Springer

Professor Hung T. Nguyen
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003-8001
USA
E-mail: hunguyen@nmsu.edu

Professor Berlin Wu
Department of Mathematical Sciences
National Chengchi University
Taipei 11605, Taiwan
E-mail: berlin@nccu.edu.tw

Library of Congress Control Number: 2005938944

ISSN print edition: 1434-9922
ISSN electronic edition: 1860-0808
ISBN-10 3-540-31695-7 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-31695-4 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springer.com

Printed in The Netherlands

Typesetting: by the authors and TechBooks using a Springer LaTeX macro package

Printed on acid-free paper SPIN: 11353492 89/TechBooks 5 4 3 2 1 0

To Pattama and Ching-Min

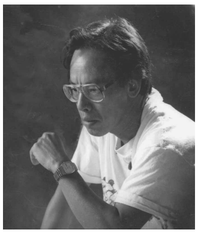
Preface

This monograph aims at laying down a rigorous framework for statistical analysis of fuzzy data. By fuzzy data we mean imprecise data which are recorded linguistically, i.e. expressed in some natural language as opposed to precise numerical measurements. Clearly, this type of data is more complex and general than set-valued observations (in, say, coarse data) which generalize data in statistical multivariate analysis.

Fuzzy data need to be modeled mathematically before they can be subject to analysis. In this monograph, we will model fuzzy data as fuzzy sets in the sense of Zadeh. Postulating that those data are generated by random mechanisms, we will proceed to formulate the basic concept of random fuzzy sets as bona fide random elements in appropriate metric spaces. With this mathematical model for populations, we are entirely in the framework of standard statistical analysis. With this goal, this monograph is in fact more oriented towards probabilistic foundations than statistical procedures. We formulate, however, basic problems in statistics such as estimation, testing and prediction. We will also illustrate statistical methods for analyzing fuzzy data through some studies.

It is our hope that this monograph will serve as a solid starting point for analyzing fuzzy data and thus enlarging the domain of applicability of the field of statistics.

We thank Janusz Kacprzyk, editor of the series, for encouraging us to write this monograph. Special thanks are due to Ding Feng for allowing us to use his recent thesis on convergence of random sets. We thank our families for their love and support.



Hung T. Nguyen

December 2005 *Berlin Wu*

Contents

Chapter 1 Introduction

First like fuzzy logics are logics with fuzzy concepts, by fuzzy statistics we mean statistics with fuzzy data. Data are fuzzy when they are expressed in our natural language. For example, the linguistic value "young" for the age of Tony in "Tony is young" is a fuzzy concept. We use natural language to describe phenomena when precise measurements are not available. The fuzzy concept "young" is intended to describe Tony's age from our perception. As such it should "contain" the true age of Tony. However, the boundary of a set which supposes to "contain" Tony's true age is not sharply defined. We will call this vagueness in the meaning of "young" fuzziness. We take fuzzy concepts in our nature language as primitives. We use our natural language to impart knowledge and information.

This monograph is about the topic of fuzzy statistics. Thus let us address three basic questions. *What* is fuzzy statistics? *Why* do we need fuzzy statistics? *How* do we intend to carry out statistical analysis of fuzzy data?

1.1 What is Fuzzy Statistics?

As stated above , fuzzy statistics is statistics of fuzzy data , i.e. making inference from data which contain fuzzy concepts such as estimation , testing of hypotheses and prediction, just like in standard statistical science. The closest situation to fuzzy data is set-valued observations in which data consist of sets rather than points of some domain of interest. Set-valued observations can arise in different contexts. For example, realizations of a point process are sets (which are locally finite). Here, we are interested in the observed sets themselves, i.e. the underlying population is a *point process*.

In *coarse data* analysis, such as in biostatistics, the situation is this. The variable of interest X (population) is point-valued. However, X is unobservable, and instead we observe some *coarsening* of X, i.e. a random set S containing X with probability one. Here, the population of interest is

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 1–4 (2006)
www.springerlink.com

X, and we need to use the *imprecise data* form the *coarsening population S* to carry out valid inductive logic about X. Set-valued observations are the most general form of imprecise data in standard statistics, as they generalize imprecise data such as censored data, missing data, grouped data. In this simple form of imprecise data, data are subsets of some set. Subsets have clearly defined boundaries. Fuzzy concepts are a step beyond concepts which can be described with ordinary set theory.

1.2 Why Do We Need Fuzzy Statistics?

We rely on data, either by performing experiments or by observing phenomena, to discover knowledge. This is the essence of empirical science. The forms of data obtained depend clearly on problems of interest. When data are imprecise to the extend that they can be only described linguistically, we are facing fuzzy data. Risk assessments, or interest rates, say, by experts, are classified as “low”, “moderate”, and “high”; *interest rates as perception-based information* or behavior of dynamical systems are, in general, described in natural language; concept in economics and marketing such as “*fidelity of costumers*”, “*effectiveness of promotions*”, are intrinsically fuzzy in nature. Of course, it is possible to reduce fuzzy concepts which can be described by set theory, say, by putting thresholds.

However, in one hand, whether such procedure will result in some significant loss of information, or, on the other hand, it is not appropriate to do so in view of subjectivity. This last point is similar to the concept of p-value in testing of statistical hypotheses as opposed to significant tests at pre-determined levels of rejection. In any case, there exist important cases in real-world applications in which we have to deal with fuzzy data.

As such, it seems obvious that we need to extend standard statistical science to cope with this new type of data. In doing so, we enlarge the domain of applicability of statistical science. In the foreword to the book of Matheron [51], which lays down the mathematical foundations for studying set-valued observations (as a new type of data), G. Watson wrote “Modern statistics might be defined as the applications of computers and mathematics to data analysis. It must grow as new types of data are considered and as computing technology advances". Due to the general form of coarse data and artificial intelligent technologies, we are facing fuzzy data as a new type of data, and as such, within Watson's vision of statistics, the task of theoretical statistics is to propose appropriate frameworks to carry out statistical inference with this new type of data.

1.3 How to Carry Out Statistical Inference with Fuzzy Data?

As pointed out by Frechet [25], nature and technology present various forms of random data. As such, a most general and appropriate framework for statistical analysis is needed. Traditional statistical data are vectors in Euclidean spaces or curves in function spaces. The corresponding sample spaces are metrizable, and metric spaces are appropriate mathematical structures for sample spaces. For set-valued observations (see e.g. [81], [24], and [61]), the sample space of closed sets in a Hausdorff locally compact, and second countable topological space is metrizable, when equipped with the hit-or-miss topology (see [51]). Data as sets are viewed as realizations of *random elements* with values in the space of sets.

The situation for fuzzy data is similar. To arrive at a framework, we need first two basic steps:

(i) Modeling fuzzy data
(ii) Specifying metric spaces of fuzzy data

We are going to use Zadeh's fuzzy sets [117], see also [62] to model mathematically fuzzy concepts. The essentials of fuzzy sets will be presented in Chapter 3. Chapter 2 will be devoted to a simple form of *imprecise data*, namely *random sets* whose generalization to *random fuzzy sets* will be the topic of Chapter 4. Basic aspects of statistical inference with fuzzy data will be discussed in Chapter 5 where statistics with set-valued observations will provide the foundations for extensions. The Chapter 6 provides some asymptotical analysis and aspects of decision theory related to fuzzy data. The second part of the monograph, namely Chapters 7-9, is devoted to illustrations of applications with fuzzy data.

Chapter 2 Set-valued Data

Since fuzzy sets are generalizations of ordinary sets, we present in this Chapter the essentials of random set theory for statistics. This material is useful for considering random fuzzy sets later.

2.1 Random Closed Sets

Set-valued mappings are formulated as random elements as follows.

Let $(\Omega, \mathcal{A}, \mathcal{P})$ be a probability space. Let (V, v) be a measurable space. Then a *random element* with values in V is a map $X : \Omega \to V$ which is $\mathcal{A}$-v-measurable, i.e. $X^{-1}(v) \subseteq \mathcal{A}$. The probability law of X is the probability measure P_X on v defined by $P_X = PX^{-1}$.

We are interested in the case where V is a collection of subsets of some set U. In this case, the random element X is called a *random set*. Thus, a random set is a random element whose values are sets. When U is a *finite set*, V is its power set 2^U, and v is the σ-field on 2^U consisting of all subsets of 2^U, i.e. the power set of 2^U. Its probability law P_X is then determined simply from its *distribution function*

$$F: 2^U \to [0,1],$$

where $F(A) = P(X \subseteq A)$, or equivalently, by its capacity functional

$$T: 2^U \to [0,1],$$

where $T(A) = P(X \cap A \neq \phi) = 1\text{-}F(A^c)$, where A^c denotes the set complement of A, i.e. $A^c = U\text{-}A$.

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 5–12 (2006)
www.springerlink.com

Viewing U as a topological space with its discrete topology, any *finite random set X* is a random closed set, i.e. taking closed subsets of U as values.

Random intervals $[X,Y]$ on the real line $\mathcal{R}$, where X, Y are real-valued random variables, are examples of random closed sets on Euclidean spaces.

Point processes on R^d are also random closed sets. For example, for Poisson process on R^+, the collection (T_n $n \geq 1$) of successive arrival times is a (almost sure) random set on R^+. Indeed, for any $t \in R^+$, $EN_t = rt < +\infty$, so that $N_t < \infty$ a.s., where N denotes the number of events occurring during $[0, t]$, and r is its arrival rate. Thus, except on a set of probability zero, the subset $\{T_n(\omega), n \geq 1\}$ of R^+ is locally finite in R^+ in the sense that each bounded subset B of R^+ contains only a finite number of the point $T_n(\omega)$. This is so because, for given B, we take t so that $B \subseteq [0,t]$, and hence # $(\{T_n(\omega), n \geq 1\} \cap B) \leq N_t(\omega) < \infty$, where # denotes cardinality.

As such, $\{T_n(\omega), n \geq 1\}$ is necessarily closed. Indeed, let $x \notin \{T_j(\omega), n \geq 1\}$ and $B=\{y \in R^+ : |x-y| \leq \varepsilon\}$. If $\{T_j(\omega), n \geq 1\}$ which are in B, then $B^o \bigcap \{T_j(\omega), j = 1,2,\ldots,m\}^c$ (B^o denotes the interior of B) is an open neighborhood of x which does not contain only element of $\{T_n(\omega), n \geq 1\}$, and hence $\{T_j(\omega), n \geq 1\}$ is equal to its closure.

Next, observe that a Poisson process can be defined as a *point measure*. For background on point process, see a text like [71]. This allows the extension of Poisson process to arbitrary spaces, such as $\mathcal{R}^d$, as a (random) point measure on the Borel $\sigma-$ *field* $\mathcal{B}(\mathcal{R}^d)$ of $\mathcal{R}^d$. Point processes model random distributions of *points* in the space under consideration. An extension of this model to the case of random distributions of *subsets* is useful in applications see e.g. [14].

Specifically, we need to consider Poisson model in the space $\mathcal{F}$-$\{\phi\}$ of non-empty closed subsets of $\mathcal{R}^d$. A point process is a random point measure on $\mathcal{F}$-$\{\phi\}$. We need to topologize $\mathcal{F}$. It can be shown that, like the case of Poisson process on $\mathcal{R}^d$, realizations of point process on $\mathcal{F}$-$\{\phi\}$ are closed subsets of $\mathcal{R}^d$.

Random elements with values in the space of closed sets of $\mathcal{R}^d$ (or more generally, in a Hausdorff topological space, locally compact and second countable) seem general enough for applications. Their theory is firmly established in [51].

In the following, $\mathcal{F}$, $\mathcal{K}$, $\mathcal{G}$ denote the spaces of closed, compact and open subsets of R^d , respectively. We are going to topologize $\mathcal{F}$.

Let

$$\mathcal{F}_A = \{ F \in \mathcal{F}: F \cap A \neq \phi \}$$

$$\mathcal{F}^A = \{ F \in \mathcal{F}: F \cap A = \phi \}$$

$$\mathcal{F}^K_{G_1, G_2, \ldots, G_n,} = \mathcal{F}^K \cap \mathcal{F}_{G_1} \cap \ldots \cap \mathcal{F}_{G_n}$$

where $n = 0$, $\mathcal{F}^K_{G_1, G_2, \ldots, G_n,}$ means $\mathcal{F}^K$.

It can be check that $\mathcal{B} = \{\mathcal{F}^K_{G_1, G_2, \ldots, G_n,}, n \geq 0, K \in \mathcal{K}, G \in \mathcal{G}\}$ is qualified as a base for a topology on $\mathcal{F}$. The associated topology of τ is called the *hit-or-miss* topology of $\mathcal{F}$. The associated Borel $\sigma -$ *field* on $\mathcal{F}$ is denoted as $\mathcal{B}(\mathcal{F})$.

Definition 2.1 *Let (Ω, $\mathcal{A}$, $\mathcal{P}$) be a probability space. A random closed set on R^d is a map $X: \Omega \to \mathcal{F}$ which is $\mathcal{A}$-$\mathcal{B}$ ($\mathcal{F}$)-measurable.*

The topological space $\mathcal{F}$ is Hausdorff compact and second countable, and as such $\mathcal{F}$ is metrizable. Thus, random closed sets are random elements with values in a metric space. Since probability laws of random closed sets are probability measures on Borel σ-fields of metric spaces, we are in the standard framework of weak convergence of probability measures for the study of limit theorems for large sample statistics.

Like random vectors in $\mathcal{R}^d$, random closed sets on $\mathcal{R}^d$ can be characterized by simple objects, namely their *capacity functionals*. Specifically, if X is a random vector, i.e. $X: \Omega \to \mathcal{R}^d$ which is $\mathcal{A}$-$\mathcal{B}(\mathcal{R}^d)$-measurable, then $S=\{X\}$ is a random closed set.

Let

$$T: \mathcal{K} \to [0,1]$$

$$T(K) = P(\{X\} \cap K = \phi)$$

$$= P(X \in K)$$

Now, since the probability law P_X of X is tight (since R^d is a separable, complete metric space), we have:

$$\forall A \in \mathcal{B}(R^d)$$

$$P_X(A) = \sup\{P_X(K): K \in \mathcal{K}, K \subseteq A\}$$

Thus, P_X is completely characterized by the capacity functional T.

Next, observe that T satisfies the following.

(a) $0 \le T \le 1$, $T(\phi) = 0$

(b) If $K_n \downarrow K$ in $\mathcal{K}$ (i.e. K_n is a decreasing sequence of compact sets such that $\bigcap_{n \ge 1} K_n = K$), then $T(K_n) \downarrow T(K)$

(c) T is monotone increasing on $\mathcal{K}$ (i.e. $K_1 \subseteq K_2 \Rightarrow T(K_1) \le T(K_2)$), and for $n \ge 2$, and $K_1, \ldots, K_n \in \mathcal{K}$,

$$T(\bigcap_{i=1}^n K_i) \le \sum_{\phi \ne I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} T(\bigcup_{i \in I} A_i).$$

This property c is referred to as the *alternating of infinite order* property of T.

If S: $\Omega \to \mathcal{F}$ is an arbitrary random closed set, then *its capacity functional*

$$T : \mathcal{K} \to [0,1]$$

$$T(K) = P(S \cap K \neq \phi)$$

still satisfies the above properties (a), (b), and (c).

The remarkable Choquet theorem is that these properties determine completely the probability law P_X on S on $\mathcal{B}(\mathcal{F})$. This is the counter-part of Lebesgue-Stieltjes theorem which leads to the axiomatic definition of distribution functions of random vectors.

Specifically, a set function T on $\mathcal{K}$ is called a *capacity functional* if it satisfies the properties (a), (b), and (c) above. The Choquet theorem is this (see [1] for a proof).

Choquet Theorem. *Let $T : \mathcal{K} \to [0,1]$ be a capacity functional. Then there exists a unique probability measure Q on $\mathcal{B}(\mathcal{F})$ such that: $\forall K \in \mathcal{K}$, $T(K)=Q(\mathcal{F}_K)$.*

Choquet Theorem allows us to study probability measures on $\mathcal{B}(\mathcal{F})$ (probability laws of random closet sets) at a simple level, namely studying set functions on $\mathcal{K}$.

2.2 Coarse Data

As we will see in subsequent Chapters, when variables (or attributes) of interest in statistical problems are qualitative in nature, i.e., expressed in linguistic terms , the problem of *meaning representation* is crucial. In standard statistics, the situation of *coarse data* seems to be the closest case to meaning representation of imprecise data.

Coarse data refer to data with low quality. We run into this type of data in situations such as missing data, censor data (say, in survival analysis), and grouped data. It is the difficulty in observing accurate data that leads to coarse data. This might happen in performing random experiments as well as observing random phenomena in nature.

Specifically, suppose the random vector X of interest cannot be observed with accuracy. The statistician then tries to extract as much as possible useful information about the values of X, say, by localizing them in "observable" subsets of R^d. In other words, when we cannot subsets of X with accuracy, we *coarsen* its space of values, i.e., replacing R^d with some collection of subsets of it, called a *coarsening scheme*.

For example, suppose the random sample $X_1, X_2, ..., X_n$ from X cannot be observed, but each X_i can be located in one of the elements A_j of a measurable partition $\{A_j$, j=1,2,…m$\}$ of $\mathcal{R}^d$. Thus the coarse data associated with $X_1, X_2, ..., X_n$ is $A_1, A_2, ..., A_n$ (where $X_i \in A_i$, almost sure). Put it differently, let $A_1, A_2, ..., A_n$ be a partition of R^d. Suppose each value of X can be located in exactly one of the A_i's. Then the problem of coarse data is modeled by a (finite) *random set S* taking values in $\{A_1, A_2, ..., A_n\}$ with

$$P(S = A_i) = P(X \in A_i),\ i=1, 2, \ldots, n.$$

It is clear that the relation between the variable X and its coarsening scheme S is that X is an almost sure (*as*) *selector* of S, i.e., $P(X \in S) = 1$.

Let f be the unknown probability density of X. If the random sample $X_1, X_2, ..., X_n$ from X are observable, then problem of non-parametric estimation of f can be handled either by the Kernel method or by Fourier method. The set-valued observations considered in this example is an extension of multivariate statistics. For example, suppose f is parametrized as $f(X \mid \theta))$ for $\theta \in \Theta$. Then the likelihood function, based on $S_1, S_2, ..., S_n$, (where $X_i \in S_i$, i=1,2,…,n) which form a random sample from the random set S, is

$$L(\theta \mid S_1, S_2, ..., S_n) = \prod_{i=1}^{n} \int_{S_i} f(x \mid \theta) dx$$

Thus the maximum likelihood estimator of θ can be computed using the observable $S_1, S_2, ..., S_n$. Of course, to investigate large sample properties of the maximum, likelihood estimator, we need to develop distribution theory and limit theorems for random sets. For more information on coarse data, see e.g. [34], [28].

Let P_X denote the probability law of *X* on *U* (e.g. *U* is finite set or a locally compact, Hausdorff, second countable topological space like R^d). Since *X* is a.s. selector of *S*, P_X is dominated by the capacity function T_S (or simple *T*) of *S*, i.e.

$$\forall A \in \mathcal{B}(U),\ P_X(A) \le T(A),$$

where $\mathcal{B}(U)$ is the Borel $\sigma -$ *field* of subsets of *U*.
Note that the capacity functional *K* is extended from $K(R^d)$ to $B(R^d)$ as:

$$\forall A \in B(R^d),\ T(A) = \sup\ \{T(K) : K \in \mathcal{K},\ K \subseteq A\}.$$

But this condition characterizes the *core* of *T*, denotes as $\zeta(T)$, i.e. the set of all (Borel) probability measures on $\mathcal{B}(U)$ which as dominated by *T*, in view of Noberg's theorem [66]. In the context of statistics, the parameter space for P_X is $\zeta(T)$. Of course, if *T* is a capacity functional, then $\zeta(T) \neq \phi$. For a comprehensive, but introductory, to statistical inference based on random set observations, see [61].

Using coarsening schemes to model imprecise data is a practice in many fields including biostatistics. It turns out that coarsening schemes in their most general cases are the way humans gather information leading to remarkable results in control and decision, say, this is usually attributed to *intelligence*. We are talking about *perception-based information* (Zadeh, [121]).

However, the remarkable capability of humans is based on *fuzzy data*, i.e. on descriptions of observed facts by natural languages in which fuzzy concepts are essential. As we will see, fuzzy data are more general than set-valued data, where *meaning representation* becomes a fundamental issue. In order to address the problem of meaning representation of observed data, we need first to develop a mathematical theory generalizing ordinary (or crisp) set theory, and that is the topic of our next chapter.

Chapter 3 Modeling of Fuzzy Data

Fuzzy data are imprecise data obtained from measurements, perception or by interviewing people. Typically, those data are expressed in linguistic terms (in qualitative form). For example, *"Tony is young"*, is a useful information, and yet it is not quite clear how we could model it mathematically for processing. The adjective *young* is intended to give an elastic constraint on the unobservable variable X (age of Tony), the range of X is [0, 100], say.

The fuzzy concept A = *young* is not an ordinary (crisp) subset of U=[0,100] since it does not have sharply defined boundaries. Of course, when facing a problem of this kind, i.e. how to interprete the meaning of a fuzzy concept, we can put thresholds and come up with approximations, and then just use standard set theory. For example, with rationale for rejecting a null hypothesis in statistics we use the α-level of significance. We used to explain to students that if the null hypothesis is true, then it is *unlikely* that we will get the observed data that we already got, and hence we should reject the null hypothesis. The level of significance α (subjective) is the threshold for interpreting the fuzzy concept *unlikely*. The concept of *p*-value is a little more flexible in which it exhibits subjectivity. Also, in statistical practices, we face always questions of the form *how large is large*? such as in the determination of sample sizes or using χ^2-statistics in testing of categorical data.

The question of modeling of fuzzy data (concepts) is this: How to represent mathematically fuzzy concepts, such as *young age*?

Following Zadeh [117], we are going to use fuzzy theory to model fuzzy data. Fuzzy sets are mathematical objects, generalizing crisp sets, which are designed to model the vagueness present in our natural language when we observe or describe phenomena that do not have sharply depend boundaries. By doing so, we *enlarge* the class of coarse data for scientific

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 13–34 (2006)
www.springerlink.com

studies. As a by product, we enlarge the domain of applicability of statistical science.

While the theory of fuzzy sets and logics is well-understood by now, after 40 years of developments with numerous successful applications in a variety of fields, see e.g. [62], we choose to present next the essentials of the theory, especially, for statisticians who are not familiar with fuzzy set theory.

3.1 Fuzzy Sets

Let U be a set. A subset of U is characterized by its *indicator function* (denoted as A), i.e

$$I_A : U \to \{0,1\}$$

$$I_A(u) = \begin{cases} 1 & if\ u \in A \\ 0 & if\ u \notin A \end{cases}$$

The indicator function I_A of a subset A *of U is a membership function* in the sense that its values indicate whether a point u in U is a member of A or not. Here, there are only two degrees of membership : 1 for full membership and 0 for non-membership. The correspondence between A (as a specified collection of elements) and its membership function A is a bijection. Since A is a mathematical entity well defined, the use of I_A is often just a matter of convenience, i.e. without I_A we still can manipulate A. The situation is different for fuzzy concepts. Now that in probability theory, sets are used to describe events. An event such that A = *Tony's age is less than 40* can be described either by A=[0,40) or by its membership function:

$$I_A \quad : [0,100] \to \{0,1\}$$

$$I_A(u) = \begin{cases} 1 & \text{if } 0 \le u < 40 \\ 0 & \text{if } u \ge 40 \end{cases}$$

Now consider the fuzzy event B = *Tony is young*, or more specially B is a *fuzzy subset* of $U = [0,100]$ .Suppose that absolutely we need to define a mathematical entity to represent this fuzzy set B. By itself, B is not well-defined as a mathematical object in set theory. But as humans, we do understand B and it is a useful information to use. Clearly if fuzzy sets can be defined they should generalize crisp sets. Now in view of the bijection between sets and their membership functions, we should try to formalize the concepts and use these generalized membership functions to *define* fuzzy sets.

The concept of membership in crisp sets will be extended to accommodate for *partial membership*, i.e. with *degrees of membership* varying gradually from 0 to 1.

Definition 3.1 A *fuzzy subset* of a set U is a function from U to [0,1]

For example the fuzzy concept A = *young* is a fuzzy subset of U = [0,1], i.e. a function*, still denoted by A*, $A:U \to [0,1]$, where $A(u)$ is the degree to which the age u is considered as *young*.

Clearly the specification of the membership function A for a fuzzy concept A is the most delicate problem. But once $A(\cdot)$ is specified, it defines mathematically the fuzzy concept in question. As in the general theory of *semantic information*, subjectivity in meaning representation is unavoidable.

From the above definition, a rigorous theory of fuzzy sets can be established. Moreover, from a practical point of view, especially in statistics, how to obtain membership functions for fuzzy data is the most difficult question to answer. This question is still in debate today, e.g. [84] within the statistics community.

But that is a healthy sign since it shows that statisticians finally realize that if new types of data, such as fuzzy data, need to be taken into consideration to enlarge the domain of applicability of statistical science, then fuzzy set modeling of fuzzy data seems to be a bona fide mathematical approach. Note that statisticians are also considered with either quantifying linguistic probabilities, e.g. [55], or meaning the vague meaning of probability expressions, e.g. [97].

There are several approaches to the problem of determination of membership functions of fuzzy data. Membership functions can be subjectively assigned by experts, say. Degrees of membership can be obtained statistically by using relative frequencies of yes/no questions. The following formal consideration between fuzzy sets and random sets (Goodman [29]) suggests a way to model membership functions.

Let $f: U \rightarrow [0,1]$. Then f is the *covering function* of the random set S which is the randomized α-level set of f. Specifically, let α be a random variable, defined on a probability space (Ω, A, P), uniformly distributed on [0,1]. Consider the random set S: $\Omega \rightarrow 2^{\Omega}$ (the power set of U).

$$S(w) = \{u \in U, f(u) \geq \alpha(w)\}$$

Then, $\forall u \in U$,

$$P(u \in S) = P(w: \alpha(u) \leq f(u)) = f(u)$$

Thus, given P and S, a membership function f can be obtained as $f(u) = P(u \in S)$. In practice, it remains to interprete P and S. The following extension is due to Orlowski [67]. Consider the following example in subjective evaluations (multi-criteria decision problems). Suppose that you wish to rank the houses for sale on the market in order to buy the most appropriate one. Of course you want to buy a *good* house. Now A =*good house* is a fuzzy concept. If you have the membership function $A(.)$ of A, then among the set of houses H, of course you will choose the house h such that $A(h) = \max_{g \in H} A(g)$, in the sense that its meaning comes from a collection of its attributes (or criteria) C such as *good location, good price*, *good physical condition*. In other words, this list of relevant criteria provides the meaning to the fuzzy concept A. For each $c \in C$, consider the set of houses which have the criteria c (i.e. satisfy the property c), i.e. consider the multi-valued map

$$S: C \to 2^H \text{ where}$$

$$S(c) = \{ h \in H : h \text{ satisfies } c\}$$

For each $h \in H$, we are interested in the set of criteria for with h satisfies, i.e

$$\{ c \in C : h \textit{ satisfies } \mathrm{c}\} = \{ c \in C : h \in S(c) \}.$$

Given a house h which satisfies a collection of criteria, we ask experts to express their judgments on this house. Specifically, what is the degree to which you consider this house as a good house? i.e. what is $A(h)$?

Now clearly, experts express their opinion in terms of a *set function* π

$$\pi : 2^C \to [0,1]$$

so that $A(h) = \pi \{ c \in C : h \in S(c) \}$.

Here $S : C \to 2^H$ is a set function. From a common sense viewpoint, π should at least satisfy the axioms:

$$\pi(\phi) = 0, \text{ and for } a,b \subseteq C, \text{ if } a \subseteq b \text{ then } \pi(a) \leq \pi(b),$$

i.e. π is a set function which is monotone increasing.

Such set functions are familiar in many field of analysis, such as in economics (as coalitional games, see e.g. [34]), robust statistics (e.g. [37]), random sets statistics (e.g. [61]) and intelligent technologies (as fuzzy measures, e.g. [62]).

The set function π (which need not be additive) is a model for *subjective evaluation processes* from which membership functions of fuzzy data can be proposed.

Membership functions, as functions from U to [0,1] are models for meaning representation of fuzzy concepts. For example, the fuzzy concept of B = *high income* in a population is modeled as a function $B{:}U \rightarrow [0,1]$, where $B(u)$ represents the degree to which the income u is compatible with the meaning *high*. The value of the $B(u)$'s can be obtained, for example, by a random survey. Suppose for u = \$50,000, 30 people among 505 classify u as high, then B(50,000) = 30/55. But 30/55 does not mean that the population of an (annual) income of \$50,000 is high. It is not the probability of having a high income. In other words, probabilities should not be confused with degrees of membership in a fuzzy set. The message is this: Membership functions are not probability distributions. Fuzziness and randomness are two distinct concepts of uncertainty.

Roughly speaking, randomness is about the occurrences of events, whereas fuzziness is about the imprecision of meaning. Note that *high* income is a fuzzy event. Thus if the income is a random variable, defined on (Ω, A, P) with values in R^+, then a fuzzy event B is not an element of the σ-field A of ordinary events.

If we wish to talk about the probability of having a high income, where *high income* is modeled by a fuzzy subset B of R^+, then as Zadeh proposed [118], the probability of the fuzzy event B can be taken as a natural extension of probability measure from events to expectation of random variables. Specifically, if the membership function B *is measeuable, so that* B is viewed as a random variable, its expectation

$$\mathrm{E}(B) = \int_{\Omega} B(w)dP(w)$$

is by definition, the probability of obtaining the fuzzy event B. Of course, when B is crisp (i.e. non fuzzy , such as *incomes above \$40,000*) B *is* reduced to an indicator function, and the above formula is the probability of event B.

An interesting relation between fuzzy sets and ordinary sets is this. For $\alpha \in [0,1]$, let $A_\alpha(f)$ or simply A_α, when no possible confusion, be the α *-level set* of a membership function f: $U \to [0,1[$, where

$$A_\alpha = \{u \in U : f(u) \geq \alpha\}$$

Then, $\forall u \in U$, we have

$$f(u) = \int_0^1 A_\alpha(u) d\alpha$$

Where we write $A_\alpha(\cdot)$ for the indicator function of the set A_α.

3.2 Fuzzy Logics

Fuzzy logics are logics with fuzzy sets. For ordinary sets, there is one logic, the usual one we use in mathematical manipulations of sets via operators (logical connectives) of union $\bigcup$ operator, intersection $\bigcap$ and negation (set complement $(\cdot)^c$) and their implied implications / e.g. material implication $\Rightarrow$. As we will see, there is possibly an infinite number of logics for fuzzy sets. Do not *interprete* a fuzzy logic as a logic which is fuzzy, i.e. not a logic at all!

But why we need logics?! In statistics, two- valued logic (Boolean logic) is used within mathematics, and we talk rather about inductive logic for statistical inference procedures. This is exemplified from the very beginning of statistics, namely in survey sampling in which to make inductive logic valid, we introduce man-made random mechanisms into sampling designs. This process exposes for the very first time the concept of random sets in statistics since a probability sampling design is nothing else than a random set on a finite population, see e.g. [32].

In the field of mathematical logic, Boolean logic is only a special one. Fuzzy logics belong to multi-valued logics. Here we need to understand fuzzy logic in two different senses. In one sense, fuzzy logic refers to the extension of standard logic with truth values in the two-element Boolean algebra ($\{0,1\}$, $\bigcup, \bigcap$, $(\cdot)^c$, 0,1) to the case where they are the *Kleene* algebra ([0,1], $\bigcup, \bigcap$, $(\cdot)^c$, 0,1). For this investigation, interested readers are refered to [62]. Here, for statistics, we investigate fuzzy logic for the use of fuzzy sets (modeling fuzzy data) in the representation and manipulation of vague information for the purpose of statistical inference such as making decision or taking actions.

Recall that to have common notations for ordinary and fuzzy sets, we write A as a subset (fuzzy or not) and use the same notation A for its membership function. In terms of their indicator functions, the operations among ordinary sets are written as:

$$(A \bigcup B)(x) = \max(A(x), B(x))$$

$$(A \bigcap B)(x) = \min(A(x), B(x))$$

$$A^c(x) = 1 - A(x).$$

Consider first the connective *and (*intersection). The truth table of *and* is the map $\wedge : \{0,1\} \times \{0,1\} \to \{0,1\}$

B \ A	0	1
0	0	0
1	0	1

Where A and B are fuzzy sets of same set U, the extension of *and* is based on the extension of $\wedge$ to a map from $[0,1] \times [0,1] \to [0,1]$, generalizing $\wedge$. One obvious extension is $x \wedge y = \min\{x,y\}$, for $x,y \in [0,1]$. But there are many others. By examining basic logical properties of a logical *and* operator, we arrive at the following most general definition for the *and* connective for fuzzy logics.

Note that each extension of $\wedge$ gives rise to an associated fuzzy logic. This general definition is inspired from the theory of probabilistic metric spaces [23], where in the concept of *t-norms,* t stands for *triangular*. Do not confuse *t-norms* with p-norms *in* $L^p(\Omega, A, P)$ spaces!

The concept of *t-norms* in probabilistic metric spaces comes from the concept of copulas in statistics. The concept of copulas which is a Latin word meaning a link, appeared in probability literature since the work of Sklar in 1959 (see e.g. [82]) related to Frechet's problem on distributions with given marginals. There are functions which join multivariate distribution functions to their marginals. Specially, a *d-copula* is a function

$$C\colon [0,1] \xrightarrow{d} [0,1], \text{ such that}$$

(a) $\forall x_j \in [0,1]$, $C(1,\ldots,1,x_j,1,\ldots 1) = x_j$

(b) For any [a,b] = $\prod_{i=1}^{n}[a_i.b_i] \subseteq [0,1]$, $\Delta_{a,b}(C) \geq 0$

(c) C is grounded, i.e. , $C(x_1,\ldots,x_d) = 0$, i.e. for all $(x_1,\ldots,x_d) \in [0,1]$ for at least one x_i

Remark: $\Delta_{a,b}(C) = \sum_x s(x) \cdot C(x)$, where $a,b,x \in [0,1]^d$ and $s(x)$ is +1 or -1 according to whether the number of I satisfying $x_i = a_i$ is even or odd.

Example 3.1 *2*-copulas : xy, $x \wedge y$, $0 \vee (x+y-1)$.

Clearly any *t-norm* (defined below) satisfies (a) and (b). So the relation between copulas and *t-norms* hinges partly on condition (b). But commutativity and associativity are issues. In fact, a *2*-copulas is a *t-norm* if and only if it is associative, and a *t-norm* is a *2*-copulas if and only if it is *2*-increasing (i.e. satisfying (b) for $n = 2$.

Copulas are useful in statistics for models of multivariate distributions. The famous Sklar's theorem (for $\alpha = 2$) is this. Let F be a bivariate distribution function whose marginal distribution functions are H and G. Then there exists a *2*-copula C such that

$$F(x,y) = C\,(H(x), G(y)),\ \forall (x,y) \in R^2.$$

Conversely, if C is a 2-copula, and H and G are one-dimensional distribution functions, then $C(H(x),G(y))$ on R^2 is a distribution function with marginal distribution functions H and G.

In view of this historical development, fuzzy logics are related to statistics. The general extension of two-values *and* connective to [0,1] is given by a *t-norm* where,

Definition 3.2 A binary operator $\Delta : [0,1] \times [0,1] \to [0,1]$ is called a *t-norm* if it satisfies the following axioms:

For $x,\ y,\ z \in [0,1]$

(i) $1\,\Delta\, x = x$ (1 acts as an identity)

(ii) $x\,\Delta\, y = y\,\Delta\, x$ (Δ is commutative)

(iii) $x\,\Delta\,(y\,\Delta\, z) = (x\,\Delta\, y)\,\Delta\, z$

(iv) If $a \le x$ and $y \le z$ then $a\,\Delta\, y \le x\,\Delta\, z$ (Δ is increasing in each argument).

Example 3.2

(a) $x\,\Delta_0\, y = \begin{cases} x \wedge y & \text{if } x \vee y = 1 \\ 0 & \text{otherwise} \end{cases}$; where $\wedge$ =minimum, $\vee$ =maximum.

(b) $x\,\Delta_1\, y = 0 \vee (x+y-1)$

(c) $x\,\Delta_2\, y = \dfrac{xy}{2-(x+y-xy)}$

(d) $x\,\Delta_3\, y = xy$

(e) $x\,\Delta_4\, y = \dfrac{xy}{x+y-xy}$

(f) $x\,\Delta_5\, y = x \wedge y$

Thus, for example if A and B are two fuzzy subsets of R, the intersection of A and B is the fuzzy subset whose membership function is given as $x \to A(x)B(x)$ if the *t-norm* Δ_3 above is chosen.

For statistical manipulations, two main concerns arise,

(i) which *t-norm to choose?*

(ii) Are there criteria to compare *t-norm*?

For (i), first of all the apparent complexity of having so many choices for the connective *and* seems annoying for applications!

But perhaps this is the flexibility of human reasoning processes. Empirical studies have shown that fusion operators used by human experts can take different forms. The classical problem in statistics related to this issue is the problem of pooling experts' *knowledge* using statistical methods.

While in principle, there is an infinite number of possible *t-norms* for modeling conjunction (connective *and* in nature language), mathematically we can focus only on *three types* of *t-norm* (the above *t-norm* Δ_0 is not continuous), there are three types:

(a) *Idempotent t-norms.* A *t-norm* Δ is said to be idempotent if $\forall x \in [0,1],\ x\Delta x = x$. The *t-norm* $\Delta_5 = \Delta$ is the only idempotent *t-norm*

(b) *Archimedian t-norms.* A *t-norms* is *Archimedian* if it is convex (i.e. whenever $x\Delta y \le c \le u\Delta v$, then there exist a r between x and u and s between y and u such that $c = r\Delta s$), and for each *a, b*$\in$ (0,1), there is a positive integer n such that

$$a\Delta a\Delta ...\Delta a \ \ (n\ times) < b.$$

The *t-norms* $\Delta_1, \Delta_2, \Delta_3, and\ \Delta_4$ above are all Archimedian.

(c) Nilpotent *t-norms*. A *t-norm* Δ is said to be nilpotent if for $a \neq 1$, then $a\Delta a\Delta ...\Delta a$ (*n times)* = 0, for some positive integer in (where *n* depends on *a*). A *t-norm* Δ is strict if for $a \neq 0$ $a\Delta a\Delta ...\Delta a$ (*n times)* > 0 for some positive integer *n.* For example, Δ_1 is nilpotent, and Δ_3 is strict.

Via isomorphisms, we have in fact three canonical *t-norms*, namely

$$x\Delta_3 y = xy \text{ (strict Archimedian)}$$

$$x\Delta_5 y = x \wedge y \text{ (idempotent) and}$$

$$x\Delta_1 y = 0 \vee (x + y - 1) \text{ (nilpotent)}$$

For a complete analysis of the above classification of *t-norms,* see [62].

As for question (ii), we can compare *t-norms* by a *sensitivity analysis.* By associativity of *t-norms,* we can consider the sensitivity of logical connectives with respect to variations in their arguments.

For example, let *f*: $[0,1]^n \to [0,1]$. An extreme measure of sensitivity of *f* is defined as follows.

For *x*=$(x_1,...,x_n)$, $y = (y_1,...,y_n)$, and $\delta \in [0,1]$. Then the function

$$\rho_f : [0,1] \to [0,1],$$

$$\rho_f(\delta) = \max\{|f(x) - f(y)| : |x_i - y_i| \leq \delta, i = 1, ..., n\}$$

measures a kind of extreme sensitivity of *f*.

With this notion, we can proceed as in decision theory, by viewing $\rho_f(\delta)$ as the *risk* at δ of the procedure *f, so that a procedure* f is less sensitive than another procedure *g* if

$$\rho_f(\delta) \leq \rho_g(\delta)$$

for all $\delta \in [0,1]$, with strict inequality for some δ.

For example, for n = 2, and $f = \wedge$, we have $\rho_{\wedge}(\delta) = \delta$. In fact, it can be shown that $\wedge$ is the least sensitive among all continuous *t-norms*.

An alternative to the above extreme measure of sensitivity, we can consider a *measure of average sensitivity* for *t-norms.*

Let f: [a,b] $\rightarrow R$ be smooth. A measure of the sensitivity of the differentiable function f at a point $x \in$ [a,b] is $[f'(x)]^2$. Then the average sensitivity is $\frac{1}{b-a}[\int_a^b [f'(x)]^2 dx]$. More generally, for $f:[a,b]^n \rightarrow R$, the quantity

$$S(f)=\frac{1}{(b-a)^n}\int_{[a,b]^n}\sum_{i=1}^{n}(\frac{\partial f}{\partial x_i})^2 dx_1...dx_n$$

can be taken as the average sensitivity of the function f.

For example, S($\wedge$)=1, S(Δ_3)= 2/3, and S(Δ_1)=1. In fact it can be shown that Δ_3 has the smallest average sensitivity among all t-norms.

We turn now to other basic connectives in a fuzzy logic, namely connectives or, disjunction negation, and implication operators If…Then… .

Recall that in two-valued logic, the three basic connectives and, or, not are related via De Morgan's law and generate the so-called material implication operator $\Rightarrow$. Specifically, let $\wedge$, $\vee$, $(\cdot)^c$ denote the truth evaluations of conjunction (interaction), disjunction (union) and negation (complement), respectively where

$\vee$	0	1
0	0	1
1	1	1

$\wedge$	0	1
0	0	0
1	0	1

	'
0	1
1	0

And $A \Rightarrow B = A' \vee B$

$\Rightarrow$	0	1 '
0	0	1
1	1	0

In fuzzy logic, the truth set $\{0,1\}$ is replaced by the interval[0,1]. An obvious extension of the basic connectives is: for $x,y \in [0,1]$,

$$x \vee y = \max\{x, y\}$$

$$x \wedge y = \min\{x, y\}$$

$$x' = 1 - x$$

But since $x \wedge y = \min\{x, y\}$ is a *t-norm,* there are other *t-norms* for the extension of $\wedge$ for $\{0,1\}^2$ to $[0,1]^2$.

Now two-valued logic forms a De Morgan system , i.e.

$$A \bigcup B = (A' \bigcap B')'$$

$$A \bigcap B = (A' \bigcup B')'$$

so that $\vee$ is dual to $\wedge$ with respect to $(\cdot)'$, $x \vee y = (x' \wedge y')'$.

Thus, for any *t-norm* Δ, and a negation $(\cdot)'$, the disjunction of fuzzy sets is modeled by a *t-conorm* ∇ where

$$x \nabla y = (x' \Delta y')'$$

More specifically, a binary operator on [0,1] is a *t-conorm* if it satisfies

$$\begin{cases} 0 \nabla x = x \\ x \nabla y = y \nabla x \\ x \nabla (y \nabla z) = (x \nabla y) \nabla z \\ \textit{If } \ w \leq x \textit{ and } y \leq z, \textit{then } w \nabla y \leq x \nabla z \end{cases}$$

For example, with $x' = 1 - x$, the dual *t-conorm* of $x \Delta y = 0 \vee (x + y - 1)$ is $x \nabla y = 1 \wedge (x + y)$.

For negation operators, we can also consider the class of operators $\eta : [0,1] \to [0,1]$, such that:

(a) $\eta(0) = 1, \eta(1) = 0$

(b) η is non-increasing

(c) $\eta(\eta(x)) = x$

similarly, *fuzzy implication operators* are generalizations of material implication operator of two-valued logic. Specifically, a *fuzzy implication* is a map

$$\Rightarrow : [0,1] \times [0,1] \to [0,1],$$

where restriction to $\{0,1\} \times \{0,1\}$ coincides with the truth table of the material implication. Some examples are

(i) $(x \Rightarrow y) = \begin{cases} 1 & if\ x \le y \\ 0 & if\ x > y \end{cases}$

(ii) $(x \Rightarrow y) = 1 \wedge (1 - x - y)$

(iii) $(x \Rightarrow y) = y \vee (1 - x)$

Jut like *t-norms, t-conorms* and negation operators, the choice of fuzzy implications depend on the problems under consideration.

In summary, there is a variety of fuzzy logics, each of which is formed by a choice of basic connectives and implication operators.

3.3 Fuzzy Relations

An ordinary n-array relation is a subset $\mathcal{R}$ of the Cartesian product $U_1 \times U \times \cdots \times U_n$ of n sets. Thus an n-array *fuzzy relation* in a set $V = U_1 \times U \times \cdots \times U_n$ is a fuzzy subset $\mathcal{R}$ of U, i.e. R: $V \to [0,1]$.

For example, R= *much more important than* is a 2 -array fuzzy relation . Let us elaborate a little bit about this fuzzy relation in decision-making, say, by ranking alternatives using Saaty's method of Analytic Hierarchy Process (AHP), see [74]. [75].

In research and development project selections, the problem is somewhat similar to the problem of determining of a membership fuuction of a fuzzy concept. Specifically, suppose we wish to choose an option amomg a set of options (or alternatives), say choosing a house to buy among m houses. This decision problem should be carried out on the basis of a set of criteria, say *location, price and physical condition.*

The AHP ranking procedure is a linear operator (weighted average of degrees of importance of criteria). The real-world problem itself is qualitative. To obtain those degrees, the AHP proposes to conduct pairwise comparisions between criteria. For example how to assess numerically the *weight* of statement such as *criteria i is strongly more important than criteria j*. Thus , in problems such as these, fuzzy relations appear naturally and have to be handled as such.

3.4 α-Level Sets of Fuzzy Sets

The following might be somewhat familiar to statisticians. Let *f* be the unknown probability density function on R^{d-} of a random vector *X.* Given a random sample $x_1, x_2, \cdots, x_n$ from *X,* we wish to estimate *f* pointwise. This is the well-known nonparametric estimation of probability density functions in statistics. The *excess mass approach* (see e.g.[27], [28]) is an alternative to *kernel method* and orthogonal functions when qualitative information about the density *f* (such as its shape, geometric properties, of its contour clusters) is available rather than analytic information.

The density function $f: \mathcal{R}^d \to \mathcal{R}$ can be written in terms of its α-level sets $A_\alpha(f)$, or simply A_α, as

$$\forall x \in \mathcal{R}^d,\ f(x) = \int_0^\infty A_\alpha(x) d\alpha ,$$

where for $\alpha \geq 0$,

$$A_\alpha = \{x \in \mathrm{R}^d : f(x) \geq \alpha\}$$

and we write $A_\alpha(\cdot)$ for the indicator function of the set A_α.

Thus the estimation of $f(x)$ can be reduced to the estimation of the sets A_α 's. Specifically. let $A_{\alpha,n}$ ($x_1, x_2, \cdots, x_n$) be a *set –estimator* of A_α based on the random sample $x_1, x_2, \cdots, x_n$, then $f(x)$ can be estimated by the plug –in estimator

$$f_n(x) = \int_0^2 A_{\alpha,n}(x) d\alpha$$

To be rigorous, the $A_{\alpha,n}$'s should be bona fide *random closed sets* in the same sense of Matheron.

The plausible set-estimators $A_{\alpha,n}$ are derived by the *excess mass approach* in the same spirit as *Maximum Likelihood Estimators.* Indeed, for a fixed α , the target parameter is the set A_α . The qualitative information about f leads to the statistical model $A_\alpha \in \zeta$, where ζ is a specified class of subsets of $\mathcal{R}^d$, e.g. the class of *closed* convex subsets of $\mathcal{R}^d$ or ellipsoids.

Let dF denote the probability of X on $\mathcal{B}(\mathcal{R}^d)$, i.e. the Stieltjes measure associated with f. Let λ denote the Lebesgue measure on $\mathcal{B}(\mathcal{R}^d)$. Then clearly $(dF\text{-}\alpha\lambda)\, A_\alpha$ is the *excess* of the set A_α at level α .

Consider the signed measure $\varepsilon_d = (dF\text{-}\alpha\lambda)$. Then, $\forall A \in \mathcal{B}(\mathcal{R}^d)$, since $A = (A \cap A_\alpha) \cup (A \cap A_\alpha^c)$, we have

$$\varepsilon_d(A) \le \varepsilon_d(A_\alpha),$$

i.e. the α -level set A_α has the largest excess mass, at the level α , among all Borel sets A. This suggests a way to estimate A_α using the empirical counter-part of the signed measure $dF\text{-}\alpha\lambda$.

Let dF_n denote the empirical measure associated with the random sample $x_1, x_2, \cdots, x_n$, i.e.

$$dF_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} ,$$

with δ_x being the Dirac measure at $x \in \mathcal{R}^d$. Then the empirical excess mass, at the level α, of $A \in \mathcal{B}(\mathrm{R}^d)$, is $\varepsilon_{\alpha,n}(A) = (dF_n - \alpha\lambda)(A)$.

Therefore, it is natural to hope that *good* estimators of A_α can be obtained by maximizing $\varepsilon_{\alpha,n}(A)$ over ζ. However, note that this optimization problem needs special attention.: the variable is neither a vector nor a function but a set! This type of optimization of set functions occurs in many areas of applied mathematics, such as shape optimization.

The above situation is somewhat similar to *fuzzy analysis* in which the concept of α -*level sets* of membership functions of fuzzy data is useful.

Let A be a fuzzy subset of a set U. Then $A_\alpha = \{u \in U, A(u) \geq \alpha\}$, $\alpha \in [0,1]$. Clearly the A_α's characterize A, i.e. two fuzzy sets A and B are identical if and only if $A_\alpha = B_\alpha$, $\forall \alpha \in [0,1]$.

Now let U, V be two sets and $f: U \to V$. In studying fuzzy data, we usually need to extend f from $F(U)$ *to* $F(V)$, where $F(U)$ denotes the space of all fuzzy subsets of U.

First, for an ordinary subset A of U, $f(A)$ is the image of A under f, i.e. $f(A)= \{f(u)\colon u \in A\}$ which is a subset of V. By looking at their indicator functions, we have

$$F(A)(V) = \sup \{A(u){:}u{=}f(V)\}$$

This formula is extended easily to the case where A is a fuzzy subset of U. This extension is referred to as the *extension principle* in fuzzy logic literature. This principle provides a natural way to lift functions on *points* directly to fuzzy sets. For example, all operations on R are extended to fuzzy subsets: if f: $\mathrm{R}^2 \to \mathrm{R}$, then for A,B fuzzy subsets of R, $f(A,B)$ is the fuzzy subset of R whose membership function is given by:

$$f(A,B)(z) = \sup\{A(x) \wedge B(y) : f(x,y) = z\}$$

where the supremum is taken over all (x,y) such that $f(x,y) = z$. This is particularly useful in manipulating data which are fuzzy numbers, i.e. special subsets of the real line R.

A formula relating α-level sets of $f(A,B)$ to those A and B is the following.

Let $f : U \times V \to W$. Let $[f(A,B)]_\alpha$ denote the α-level sets of $f(A,B)$, where A and B are fuzzy subsets of $U,V,$ respectively. Then

$$[f(A,B)]_\alpha = f(A_\varepsilon, B_\alpha),$$

for all $\alpha \in [0,1]$ if and only if for any $w \in W$, the supremum of $\{A(u) \wedge B(v); (u,v) : f(u,v) = w\}$ is attained.

Note that $\wedge$ is a special *t-norm.* The above result can be extended to any *t-norm* Δ as follows (see [29]). First, when $\wedge$ is replaced by Δ in the extension principle, the fuzzy set $f(A,B)$ is defined as a *sup*- Δ -*convolution*, i.e. $f(A,B)$ (z)= sup{ Δ $(A(x),B(y)$: $(x,y) \in f^{-1}(z)$ }. Then $[f(A,B)]_\alpha$ = $\bigcup_{\Delta(t,s)\geq\alpha} f(A_t, B_s)$, $\forall \alpha > 0$, if and only if for all $z \in W$, $f(A,B)(z)$ is attained.

3.5 Fusion of Fuzzy Data

Suppose that in a statistical regression problem, the independent and dependent variables *X* and *Y* are not observable with accuracy, but instead of the data ($x^{(i)}, y$), *i=1,2,…,n*, consist of fuzzy data. Specifically, suppose *X*= $(X_1, X_2, ...X_d)$ be a *d*-dimensional random vector, and *Y* is a random variables, say. The generalization of an observation *X* = *a* is *X=A* where $a \in R^d$ and *A* is a fuzzy subset of R^d. For example, *A* is the fuzzy set of vector around the origin. Thus the observed data is of the form of *If …, Then…rules.*

R_i : *If* X_1 *is* A_{1i} , and X_2 *is* A_{2i} , *.and* X_d *is* A_{di}, then *Y* *is* *B ;* *i=1,2,…,n*; where the A_{ji} *'s* and B_i *'s* are fuzzy subset of *R*.

The purpose of regression analysis is to approximate *Y=g*(*X*) based on the above pairs of data. First like neural networks, see e.g [30], the above fuzzy rule base provides a *model-free* regression method for regression. This is achieved by the process consists of several steps.

Step 1. Fuzzy modeling of the linguistic labels A_{ji} *'s* and B_i *'s* in the rule base.

Step 2. Choice of connectives, such as *and* via *t-norms,* and of fuzzy implication operator *If…, then… ;* $=$, $\Rightarrow$.

Step 3. Choice of fusion operator.

Step 4. Choice of defuzzification operator.

For example, in each rule R$_l$, each value $Y = y$ has a *rule weight* of

$$\Delta\ (A_{ji}(x), j = 1,...,n)\ \Delta\ B_i(y)\ \ (A(x), B(y)).$$

Let C stand for the property *appropriate values of* Y, which is then represented by the fuzzy subset of $\mathcal{R}$, depending on the *input* $(x_1, x_2,..., x_d)$,

$$C(y) = \nabla[\Delta(A_{ji}(x); i = 1,.., n, \text{i=1,2,…,n}\}$$

where ∇ is a dual *t-norm* for Δ.

For example, let $\nabla = \vee$ and $\Delta = \wedge$, then

$$C(y) = \max[\min(A_{ji}(x); j = 1,.., d, \text{i=1,2,…,n}\}$$

The choice of Δ and ∇ depend on problems at hand or might be justified by some criteria of sensitivity analysis.

In order to obtained a single output for y (corresponding to an input $\boldsymbol{x} = (x_1, x_2,..., x_d)$ we need to transform the membership function of C into a real number $D(C)$. Such a transformation is called a *defuzzification procedure.* For example, the centroid defuzzification is a form of taking expected value:

$$D(C) = \frac{\int yC(y)dy}{\int C(y)dy}$$

All the above is a form of *model –free* regression analysis with fuzzy data. We will elaborate a little bit with the *designed methodology* as well as its rationale.

In the above, we described an inference engine based on data which form a fuzzy rule base, without a rationale. A design or inference procedure is useful only if it leads to a *good* approximation of the input-output relation. Now the true input–output $y = f(x)$ is unknown, and we have only some fuzzy information about it via the fuzzy rule base. Thus a good design is a good approximation procedure for f.

A *design methodology* consists of a choice of membership functions for the A_s' and B_s' , of logical connective Δ, ∇ , negations, etc, and a defuzzification problem D to produce an input – output map:

$$f^* : (x_1, x_2, ..., x_d) \to y^*$$

We need to investigate to what extend f^* will be a good approximation for f. Basically, this is the problem of approximation of functions. So a design methodology will lead to a good approximation for f if for any $\varepsilon >$ 0, we can find an f^* such that $\| f - f^* \| \le \varepsilon$, where $\| \cdot \|$ denote a distance between f and f^* .

Just like neural networks, it turns out that the approximation capability can be proved by calling upon the *Stone-Weierstrass Theorem* which provides design requirements for obtaining an appropriate dense subset of functions. Specifically, the *Stone-Weierstrass Theorem* is this. Let ((U, δ) be a compact metric space. Let $C(U)$ denote the space of real–valued continuous functions on U, and let $H \subseteq C(U)$. Then H is dense in $C(U)$ if H satisfies

(i) H is a subalgebra of $C(X)$, i.e. for $a \in R$, and f, g in H, we have that af, f+g, fg belong to H.

(ii) H vanishes at no point of U, i.e. for $u \in U$, there is an $h \in H$, such that $h(u) \neq 0$.

(iii) H separates points. i.e. if $u, v \in U$, then there is an $h \in H$ such that $h(x) \neq h(y)$.

Now, the design methodology described above clearly depends on the membership functions A_{ji}, B_i; the *t-norm* Δ and *t-conorm* ∇*;* and the defuzzification procedure leading to the final output y^*. Thus a *designed methodology* is a triple ($\mathcal{M}$,$\mathcal{L}$,$\mathcal{D}$), where $\mathcal{M}$ denote a class of membership functions, $\mathcal{L}$ a class of logical connectives, and $\mathcal{D}$ a defuzzification procedure. Such a triple produces the input-output map $y^*= f^*(x)$. The approximate function f^* depends on the *sample size*, i.e. the number of rules in the rule vase, and is related as f_n. The rationale or approximation capability of fusion of fuzzy data is that, under suitable conditions, on ($\mathcal{M}$,$\mathcal{L}$,$\mathcal{D}$), the class of function $\{ f_n, n \geq 1\}$ is dense in the space of continuous functions $C(K)$ defined on a compact subset K of $\mathcal{R}^d$ or $\mathcal{R}$ with respect to the sup-norm.

For example, let $\mathcal{M}$ consists of *Gaussian type membership functions*, i.e.

$$A_{ji}(x) = \alpha_{ji} e^{-(x-a_{ji})^2 / k_{ji}};$$

Δ = product *t-norm,* ∇ = its dual *t-conorm*; and centroid defuzzification procedure.

Then, for any compact set K of $\mathcal{R}^d$, the corresponding ($\mathcal{M}$,$\mathcal{L}$,$\mathcal{D}$) approximates f to any degree of accuracy, this is only an *existence theorem,* not algorithmic.

Chapter 4 Random Fuzzy Sets

Statistical models for observations which are numbers or vectors are random variables and random vectors, respectively. Similarly, if the observations are subsets of some set, then appropriate statistical models are random sets of that set. Motivated by the desire to generalize numbers to fuzzy numbers and random closed sets to fuzzy sets, we establish here a reasonable concept of random fuzzy sets which can be used in statistical inference with fuzzy data.

4.1 Back to Sampling Surveys

Since fuzzy sets generalize ordinary (crisp) sets, random fuzzy sets will generalize random sets. For the most advanced material on random sets, see [54]. Although we have recalled some basics of random closed sets in Chapter 2, we make here a return to the origin of statistics to re-emphasize the crucial role of making models in scientific studies, as well as connecting random sets to random fuzzy sets.

Suppose we wish to select a part of a finite population U with N elements ($|U| = N$) to estimate, say the population total $\sum_{u \in U} \theta(u)$, where $\theta(u)$ is the annual income of the individual u. To make this inductive logic valid, we need to be able to select a "good sample" as well as specifying the error of the estimator $\sum_{u \in S} \theta(u)$, where S is the selected sample.

To gain public acceptance, we should select the sample *S at random*. Thus, we create a random mechanism, just like a game of chance, to select samples. A sample is a subset of the population U. The random process leading to selections of samples is formally a *random set*. Samples are selected at random. For example, if we decide to select a sample of a fixed

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 35–43 (2006)
www.springerlink.com

size n, then a *probability sampling design* could be a random set who probability density function on 2^U (the power set of *U*) is

$$f : 2^U \to [0,1]$$

$$f(A) = \begin{cases} 1/\binom{N}{n} & if\ |A| = n \\ 0 & otherwise \end{cases}$$

Here a man-made randomization is introduced to make inductive logic valid. Of course, depending upon problems at hand, different probability sampling designs (i.e. different random set distributions) can be considered, see e.g. [32].

The probability of observing $x = \theta\big|_A \in X$ for some $A \subseteq U$ is $f_\theta(x) = f_\theta(\theta\big|_A) = f(A)$

Thus, if we let

$$f_\theta(x) = \begin{cases} f(A) & if\ x = \theta\big|_A \\ 0 & otherwise \end{cases},$$

then $f_\theta(x)$ is a probability density function *X* with finite suppose. Let P_θ denote the probability measure on *X* (or some $\sigma - field$, $\sigma(X)$ of *X*), i.e. $P_\theta(.) = \sum_{A \subseteq U} f(A)\delta_{(\theta|_A)}(.)$, where $\delta_{(\theta|_A)}(\cdot)$ is the Dirac measure on *X* at the point $(\theta\big|_A)$, i.e. for $B \subseteq \sigma(X)$,

$$\delta_{(\theta|_A)}(B) = \begin{cases} 1 & if\ \theta\big|_A \in B \\ 0 & otherwise \end{cases}$$

We are led to a family of probability spaces indexed by Θ, namely, $\{(X, \sigma(X), P_\theta), \theta \in \Theta\}$.

By analogy with games of chances in which outcomes are subsets of some set, such as in sampling from finite populations, we consider uncertain subsets which arise in observing phenomena as outcomes of random sets. Thus, random elements taking subsets as values appear in the very beginning of statistical science! It is interesting to observe that the random set formulation in sampling surveys sets up the general framework for statistical inference. Indeed, let U be a finite population, and let $\theta_0 : U \to R$ be our variable of interest, where, e.g., $\theta_0(u)$ denotes the annual income of the individual u. Without being able to conduct a census θ_0 is unknown, but θ_0 is located in the parameter space $\Theta = R^U = \{\theta : U \to R\}$.

Let f denote a probability sampling design (i.e. a probability density function on 2^U). Our random experiment consists of "drawing" a subset A of U according to f. Upon selecting A, we are interested in the values of θ_0 on A, i.e. the restriction $\theta_0|_A$ of θ_0 to A. Thus, we can view that the outcome is $\theta_0|_A$, and not A per se.

The sample space of our random experiment is $X = \bigcup_{A \subseteq U} R^A$.

The statistical interpretation of this statistical model is this. The true but unknown parameter of interest is θ_0 . We make an observation $x \in X$ which is drawn according to P_{θ_0} . The problem is how to make an "*educated guess*" about θ_0? Clearly, the above formulation is sampling surveys suggests the same structure for other statistical problems, namely a statistical model is a family of probability space $\{(X, \sigma(X), P_\theta, \theta \in \Theta)\}$. In standard statistical analysis, X is R or R^d , and Θ is some finitely dimensional space (parametric statistical problems).

Random variables or vectors are special cases of random closed sets. Intervals are the simplest form of imprecision about point observations (measurements). In statistics, confidence intervals (or regions in higher dimensions) are *random intervals*. Specifically, consider the parametric statistical model $\{ f(x|\theta) : x \in X \subseteq R^m, \theta \in \Theta \subseteq R^d \}$.

Suppose our parameter of interest is $\varphi|\theta$. Give a random sample $x_1, x_2, ..., x_n$ from the population X (with true, but unknown $f(x|\theta_0)$, the essence of confidence region estimator is to find a random set $S(x_1,...,x_n)$ containing $\varphi(\theta_0)$ with high probability, e.g., $\forall \theta \in \Theta$, $P_\theta\{\varphi(\theta) \in S(x_1,...,x_n)\} \geq 1-\alpha$ for some $\alpha \in (0,1)$, and where $P_\theta(d_x) = f(x|\theta)dx$.

The situation for *fuzzy data* is somewhat similar as we proceed now.

4.2 Fuzzy Numbers

The use of fuzzy sets, as mathematical models for fuzzy data, and their associated logics spreads out to many fields in science and technology, especially in computational intelligence. Here we only have statistics in mind. As such, we are going to present basic background on structures of fuzzy sets for a rigorous framework to extend mathematical statistics to the fuzzy setting.

Imprecision in the observation of the values of random variables (or vectors) results in replacing numbers by intervals or more generally *fuzzy intervals*. Let us formulate this *coarsening scheme* in specific terms.

Let R be the set of real numbers. Elements of the space of all fuzzy subsets of R, denoted as $\widetilde{P}(R)$, are called *fuzzy quantities*. Operations on R are extended to $\widetilde{P}(R)$ by the extension principle. For example, for $A, B \in \widetilde{P}(R)$,

$$(A+B)(z) = \sup\{A(x) \wedge B(y) : (x,y) : x+y=z\}$$

$$(A \cdot B)(z) = \sup\{A(x) \wedge B(y) : (x,y) : xy=z\}$$

Note that these operations generalize interval arithmetic as well as Minkowski's operations on sets.

Fuzzy numbers are special fuzzy quantities. Intuitively, "around 0" is a fuzzy number. Thus a reasonable definition is this.

Definition 4.1. A fuzzy number is a fuzzy subset A of R such that:

(i) $A(x) = 1$ for exactly one x

(ii) The support $\{x : A(x) > 0\}$ of A is bounded.

(iii) The α *-level set* A_α of A are closed intervals

From the above definition, it is not hard to show that?

(a) Real numbers are fuzzy numbers.

(b) Fuzzy numbers A are *upper semi-continuous* (u.s.c.), i.e. A_α is closed, $\forall \alpha \in [0,1]$.

(c) Fuzzy numbers are *convex* (A fuzzy subset A of R is said to be convex if its A_α are convex, i.e. intervals).

As stated above, the use of intervals and their arithmetic is appropriate in some situations involving imprecision. When the intervals themselves are not sharply defined, we arrive at the concept of *fuzzy intervals.* Generalizing fuzzy numbers, we define a fuzzy interval A as a fuzzy subset of R satisfying

(i) A is a *normal fuzzy set* (i.e. $\exists x \in R$ such that $A(x) = 1$).

(ii) The support $\{x \in R : A(x) > 0\}$ of A is bounded.

(iii) The A_α's are closed intervals.

Thus a direct generalization of real numbers to *fuzzy data* could be a class of fuzzy subsets of $\mathcal{R}$ (or of $\mathcal{R}^d$), denoted as $\mathcal{L}(\mathcal{R})$ (or $\mathcal{L}(\mathcal{R}^d)$), where elements are *normal, convex and u.s.c. fuzzy subsets with bounded supports*.

4.3 Fuzzy Set-valued Random Elements

As stated earlier, in order to analyze fuzzy data, we need to define rigorously the concept of random fuzzy sets, i.e. random elements taking fuzzy subsets of some set U as values. That boils down to specify a class of fuzzy subsets of U and a $\sigma - field$ on that class.

Since fuzzy sets are generalizations of ordinary (crisp) sets, and fuzzy numbers closed sets of R, we will proceed to formulate the concept of random fuzzy sets generalizing the well-establish concept of random closed sets of Mathéron [51] on R^d.

Recall that the relation between crisp sets and fuzzy sets is expressed in terms of $\alpha - levels\ sets$. Let A be a fuzzy subset of a set U. For $\alpha \in [0,1]$, to $\alpha - levels\ set$ of A is the crisp subset A_α of U, where

$$A_\alpha = \{x \in U : A(x) \geq \alpha\} .$$

The membership function $A(.)$ is recovered from its A_α's via:

$$\forall x \in U, A(x) = \int_0^1 A_\alpha(x) dx .$$

Let (Ω, $\mathcal{A}$, $\mathcal{P}$) be a probability space. Suppose $S : \Omega \to \widetilde{P}(U)$. Then for each $\alpha \in [0,1]$, we have a set-valued map $S_\alpha : \Omega \to 2^U$, where $S_\alpha(\omega)$ = $[S(\omega)]_\alpha$ = $\{x \in U : S(\omega)(x) \geq \alpha\}$.

For each $S_\alpha(\omega)$ to be a closed set, say when $U = \mathrm{R}^d$, we need precisely that each fuzzy set $S(\omega)$ is upper semi-continuous (u.s.c.).

Thus, the range of S should consist of u.s.c. membership functions.

The space of fuzzy numbers $\mathcal{L}(\mathrm{R})$ (or $\mathcal{L}(\mathrm{R}^d)$) is metrizable, and random fuzzy sets can be defined as measurable maps with values in . See e.g. [17].

We specify here the concept of random fuzzy sets on Euclidean spaces R^d . For aspects of random fuzzy sets on general metric spaces with Hausdorff distance, see e.g. [51], [40].

Now observe that on a *compact, metric space* (e.g. a compact metric subspace of R^d), the topology defined by the Hausdorff metric on $\mathcal{K} \setminus \{\phi\}$ (space of non-empty compact subsets of R^d) is equivalent to the relative (hit-or-miss) topology (Chapter 2), the measurability of random closed sets is expressed simply as follows.

Let (Ω, $\mathcal{A}$, $\mathcal{P}$) be a probability space. Let $\mathcal{B}(\mathcal{F})$ denote the *Borel* σ *– field* of the space of closed subsets $\mathcal{F}(\mathcal{R}^d)$ of R^d. Then $S:\Omega\to \mathcal{F}(\mathcal{R}^d)$ is called a random closed set if for any $\Pi\in\ \mathcal{B}(\mathcal{F})$, $S^{-1}(\mathcal{F})\in\mathcal{A}$.

Similar to the situation of random vectors, it can be checked that the above measurability of S is equivalent to: for any open subset G of $\mathcal{R}^d$, $S^{-1}(\mathcal{F}_G)\in\mathcal{A}$, where $\mathcal{F}_G = \{F\in\mathcal{F}: F\cap G\neq\phi\}$. In view of all the above, we take the following approach (for more details, see [45]).

Let $\widetilde{P}_K(\mathcal{R}^d)$ denote the space of fuzzy subsets of $\mathcal{R}^d$ which are u.s.c., normal (i.e. if $f\in\widetilde{P}_K(\mathcal{R}^d)$, then $\{x\in R^d : f(x)=1\}\neq\phi$), and which have compact supports (i.e. the closure of $\{x\in\mathcal{R}^d : f(x)>0\}$) is a non-empty concept subset of $\mathcal{R}^d$).

Note that we consider *non-empty* random (closed) sets on $\mathcal{R}^d$ for practical purposes. For example, in the context of coarse data analysis, a random set S is a coarsening of some random variable X. As such, S takes non-empty subsets as values (since X is an almost selector of S).

On $\widetilde{P}_K(\mathcal{R}^d)$, a natural metric is the following. Let H denote the Hausdorff metric on $F(\mathcal{R}^d)\setminus\{\phi\}$. Then for $A,B\in\widetilde{P}_K(\mathcal{R}^d)$,

$$\widetilde{H}(A,B)=\sup_{0<\alpha\le 1} H(A_\alpha,B_\alpha)$$

is a metric (see [34]), where A_α is the α *-level set* of A.

Definition 4.2 *Let* (Ω, $\mathcal{A}$, $\mathcal{P}$) *be a probability space. Then* $S:\Omega\to\widetilde{P}_K(\mathcal{R}^d)$ *is called a random fuzzy set if for any* $\alpha\in[0,1]$, *the map*

$$S_\alpha:\Omega\to\ \mathcal{F}(\mathcal{R}^d)$$

$$S_\alpha(\omega)=[S(\omega)]_\alpha=\{x\in\ \mathcal{R}^d : S(\omega)(x)\ge\alpha\}$$

is a random closed set on $\mathcal{R}^d$.

The concept of random fuzzy set considered here is a direct generalization of random closed sets in the sense of Matheron [51]. As such, statistical inference with random fuzzy sets could benefit from the foundations of statistical inference for random closed sets (on Hausdorff locally compact and second countable topological spaces, such as $\mathcal{R}^d$), see [61].

Random fuzzy sets are population models for observations which are fuzzy sets obtained at random. Various basic properties of fuzzy data can be formulated in terms of their associated $\alpha-$*levels random sets*.

Recall that if S is a random closed set (defined on $(\Omega, \mathcal{A}, \mathcal{P})$ with values in $\mathcal{F}(\mathrm{R}^d)$, then its probability law P_S is a probability measure on $\mathcal{B}(\mathcal{F})$, i.e. for $A \in \mathcal{B}(\mathcal{F})$, $P_S(A) = P(S^{-1}(A))$, and the $\sigma-$*field* $\sigma(S)$ generated by S is $\{S^{-1}(A) : A \in \mathcal{B}(\mathcal{F})\}$, $\sigma(S)$ is a sub $\sigma-$*field* of $\mathcal{A}$.

Recall also that an (infinite) family $\{A_i : i \in I\}$ of events (i.e. elements of *a*) is said to be (mutually) *independent* if for any *finite* $J \subseteq I$,

$P(\bigcap_{j\in J} A_j) = \prod_{j\in J} P(A_j)$.

A family of sub $\sigma-$*fields* $\{a_i : i \in I\}$ of a is said to be independent if any family of events $\{A_i \in a_i : i \in I\}$ is independent. A sequence of random closed sets $\{S_n, n \geq 1\}$ is said to be independent if its associated sequence $\{\sigma(S_n), n \geq 1\}$ of sub $\sigma-$*fields* is independent in the above sense. Of course $\{S_n, n \geq 1\}$ is said to be *identically distributed* when all P_{S_n} are the same. Thus, i.i.d. samples of random closed sets are defined just like the case of random vectors.

Viewing random fuzzy sets as random elements on $\widetilde{P}_K(\mathcal{R}^d)$ with the *Borel* $\sigma-$*field* $\mathcal{B}(\widetilde{P}_K(\mathcal{R}^d)$, generated by the metric H_∞, the concept of *i.i.d. samples* of *random fuzzy sets* are defined similarly.

While random fuzzy sets are random elements in well-defined measurable spaces, the concept of *expected values* of random fuzzy sets is delicate. In fact, this situation arises for random sets as well. This is essentially due to the technical problem of integration of *set-valued functions*.

Let S be a random closed set on R^d, then the *core of S* is non-empty, where the core of S is $\zeta(S)$={X: X is an *a.s.* selector of S} i.e. $X : \Omega \to \mathrm{R}^d$, measurable and $P(X \in S) = 1$. Suppose EX exists for $X \in C(S)$, then one way to define ES is the *Aumann expectation* of S, namely

$$ES = \{EX : X \in \zeta(S)\}.$$

Thus, if S is a random fuzzy set, its expectation can be defined as a fuzzy subset of $\widetilde{P}_K(\mathcal{R}^d)$, denoted by ES, such that $\forall \alpha \in [0,1]$, $(ES)_\alpha$ is the closure of $\{EX : X \in \zeta(S_\alpha)\}$.

Chapter 5 Aspects of Statistical Inference

With the background in previous chapters, problems of statistical inference with fuzzy data should be somewhat straightforward in principle! By that we mean replacing random vectors by random fuzzy sets in all aspects of statistical inference. Of course, as in any generalization problem, this is just a guideline. Due to the nature of fuzzy data, as observations from random fuzzy sets, technical difficulties are expected in developing the theory. In fact, actual research is aiming at investigating, say, limit theorems for random fuzzy sets in order to provide rationale for large sample samples statistics with fuzzy data.

Here we choose to discuss some basic aspects of statistical inference with emphasis on "*where do fuzzy data come from*"

5.1 Fuzzy Data in Sampling Surveys

As primitive as sampling surveys in applied statistics, fuzzy data appear naturally in opinion surveys, say, questionnaires. In an opinion surveys questionnaires are sent to a random sample of individuals in a population. The data collected consists of a sample of responses to the questionnaires. As usual, the questionnaires in a sampling survey contain explicitly or implicitly variables which are either quantitative (numerical) or qualitative variables, the answers are required to be just "*yes*" or "*no*".

This practice is not just rooted in two-values logic (true or false), but perhaps, even in obvious cases such as "*Is the candidate A suitable for the specified job?*", allowing a response scale larger than just "*yes*" or "*no*", might introduce difficulties in final analysis! Realistic scales reflecting truly respondents' opinion should be either in linguistic answer or degrees to which the candidate A is suitable for the job.

Although, applied statisticians do recognize this phenomenon, the practice is that one could avoid "*fuzziness*" by using binary logic! Of course, people are entitled to do so, but the question is "*how much*

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 45–70 (2006)
www.springerlink.com

information is lost in the process?" This should be answered based on the importance of the issue involved. For example, in the risk analysis for adopting construction designs, or risk assessment of potentially critical situations, security issue is at stake.

A rough analysis based on two-values logic might be too naïve. Risk is a *linguistic variable* with values such as "*low*", "*medium*", "*high*". It is here that fuzzy set theory enters statistics to provide representation and manipulation for realistic data. Thus we can enlarge the statistical analysis of sampling surveys in which questionnaires are better designed and conclusions are more realistic.

For example, a *fuzzy proportion* of people favoring the candidate A (say in an election) is definitely more realistic to use in predication than a crisp proportion based on two-valued logic. See later chapters for concrete examples and applications in social sciences. Another typical example is in computer science: the extension of relational databases to fuzzy relational data bases allows queries to be stated in a more flexible way. For example, a standard database consisting of the variable "*name of employee*" and relation "*height*" can be extended so that a query of the form "*select tall employees*" can be retrieved.

These examples illustrate the current tendency of data analysis in extracting more information from data, as opposed to the situation where by nature, problems are fuzzy, but crisp mathematics are used to avoid the natural phenomena of fuzziness! As stated ealier, fuzzy data are not conventional in statistical practices. Perhaps this is due to the lack of a mathematical theory to analyze fuzzy data. While the statistical analysis of random set data is starting to take shape, we believe that it will not take much longer for the statistics community to embrace fuzzy statistics. Fuzzy data in sampling surveys statisticians should consider when dealing with real-world problems.

5.2 Coarse Data

Incomplete data are natural in a variety of problems and have been received attention of statisticians for quite some time (e.g. [72]) The type of incomplete data that has been studied thoroughly is missing data, in which each data value is either known or entirely unknown. But this "*yes*" or "*no*" situation is only a special case of general patterns of incomplete data. Indeed, collected data could be neither entirely missing nor perfectly present. This kind of incomplete data is termed "*coarse data*" and has received much attention more recently in statistics (e.g. [34]). A typical example of this type of "*not yes or no*" is the following situation of imprecise observations.

In performing an experiment or in observing natural phenomena, we might not be able to record correctly the values of locate them with some degree of accuracy, or more generally, locate them in some regions (subsets) of the sample space. Thus the observations are present but not known!

For example, suppose our random variable X of interest takes values in U. Suppose X cannot be observed directly but its outcomes can be located in a finite partition $\{A_1, A_2, ..., A_n\}$ of U, i.e. each $X(w)$ is known to be in some A. Such a partition is referred to as a coarsening of X. Providing a model for observing X. If we consider the set-valued map $S(w)$ = the element of the partition containing $X(w)$, then S is formally a random set with probability density

$$P(S = A_i) = P(X \in A_i),\ i=1,\dots,k$$

Thus, coarse data are actually observations of random sets which are special cases of random fuzzy sets.

The following is an example of statistical estimation in the context of coarse data. For details, see [38].

Suppose the random variable X admits a probability density function $f(x \mid \theta)$, where $\theta \in \Theta \subseteq R$. Let S be the above coarsening scheme of X. Let $S_1, S_2, ..., S_n$ be a random sample drawn from S. Then the likelihood of θ when observing $S_1, S_2, ..., S_n$ is

$$L(\theta \mid S_1, S_2, ..., S_n) = \prod_{i=1}^{n} \int_{S_i} f(x \mid \theta) dx$$

Suppose the parameter space Θ is finite. Then the maximum likelihood estimator $\hat{\theta}_n$ of θ, based on the random set observations $S_1, S_2, ..., S_n$ is unique with probability tending to one, and is weakly consistent.

This can be shown as follows:

Since Θ is finite, an $\hat{\theta}_n$ which maximizes $L(\theta \mid S_1, S_2, ..., S_n)$ over Θ exist. Let $\Theta = \{\theta_0, \theta_1, ..., \theta_n\}$ with θ_0 being the true (unknown) parameter. And let

$$P_{\Theta_j}(s_i) = \int_{S_i} f(x \mid \theta_j) dx$$

(i) Observe that $P_{\Theta}(s_i)$, i=1,…,n is an i.i.d. (independent, identically distributed) sample of a random variable *Z* which is a function of the random set *S*. Similarly,

$$\log \frac{P_\theta(S_i)}{P_{\theta_0}(S_i)}, \text{ i=1,…,n.}$$

Obviously, for these samples, the weak law of large numbers holds, e.g.

$$\frac{1}{n} \sum_{i=1}^{n} \log \frac{P_\theta(S_i)}{P_{\theta_0}(S_i)} \xrightarrow{P} E_{\theta_0}[\log \frac{P_\theta(S_i)}{P_{\theta_0}(S_i)}].$$

(ii) Since $L(\theta_0 \mid S_1, S_2, ..., S_n) > L(\theta \mid S_1, S_2, ..., S_n)$, is equivalent to

$$\frac{1}{n} \sum_{i=1}^{n} \log \frac{P_\theta(S_i)}{P_{\theta_0}(S_i)} < 0,$$

and using (i) together with the fact that –log*x* is strictly convex, we have

$$E_{\theta_0}[\log \frac{P_\theta(S)}{P_{\theta_0}(S)}] < \log E_{\theta_0}[\log \frac{P_\theta(S)}{P_{\theta_0}(S)}]$$

$$= \log[\sum_{A \in \{A_1, A_2, \ldots, A_n\}} P_{\theta_0}(S = A) \frac{P_\theta(S_i)}{P_{\theta_0}(S_i)}]$$

$$= \log[\sum_{A \in \{A_1, A_2, \ldots, A_n\}} P_\theta(S = A)]$$

$$= 0$$

Thus, for any $\theta \neq \theta_0$

$$\lim_{n \to \infty} P_{\theta_0}[L(\theta_0 \mid S_1, S_2, \ldots, S_n) > L(\theta \mid S_1, S_2, \ldots, S_n)] = 1.$$

(iii) Let B_{jn} denote the event

$$\frac{1}{n} \sum_{i=1}^{n} \log \frac{P_{\theta_j}(S_i)}{P_{\theta_0}(S_i)} < 0, \text{ j=1,…m.}$$

By (ii) we have that $P_{\theta_0}(B_{jn}) \to 1$ as $n \to \infty$, meaning that $\hat{\theta}_n$ is unique and weakly consistent.

The actual research in coarse data is focusing on modeling of set-valued observations. Specially, suppose we have a way to observe sets which contain unobserved values of *X*. The question is "*what is the relation between the random set S, giving rise to the actual coarsening scheme, and the underlying random variable X?*" Mathematically speaking, instead of looking at the "*selection problem*" i.e. given the set-valued map *S*, what are the (almost sure) selectors *X* of *S*?

Statisticians are asking the question the other way around, namely, given a measurable map *X*, what are the set-valued maps *S* which admit *X* as one of their (a.s.) selectors? Each such random set *S* will provide a model (or a coarsening scheme) for the unobservable random variable *X*. The above example is an example of the so called coarsening at random (CAR) model in the literature (e.g. [34],[28]).

The statistical inference (at a simple level, i.e. without asymptotics) of coarse data is usually based on the assumption that the coarsening scheme considered is a CAR model for the unobservable random variable under study. The advantage of this assumption is that, as illustrated from the above example, computations of estimators can be based only on observed random sets, and we can "*ignore*" the random mechanism (i.e. the distribution) of the coarsening scheme (i.e. the random set S) .

In [28] it was pointed out that the assumption of CAR model seems reasonable for the finite case, the situation for arbitrary random variables (say continuous random variables) is not clear. On the other hand, it should be noted that, even with this assumption, large sample statistics aspects require of asymptotics results such as limiting distributions of estimators (based on coarse data) which will involve distributional information about S. Here Choquet weak convergence of capacity functional of random sets seems useful. See [23]

Designing coarsening schemes for unobservable random variables are essential for extracting information. In the field of artificial intelligence, we mimic remarkable capabilities of humans to observe their environment by using qualitative concepts. In the context of coarse data, this results in extending random sets to random fuzzy sets.

For example, assessments of probabilities or likelihoods to events could be done linguistically, i.e. using a fuzzy partition of the unit interval [0, 1], such as "*low*" , "*medium*", "*high*".

5.3 Large Sample Statistics with Coarse Data

Since random fuzzy sets are generalizations of random closed sets, and are represented neatly in terms of their α-level sets the advances in statistics with fuzzy data rely obviously on those as statistics with random set observations. We discuss here a framework for statistical inference with coarse data which should form a basis for statistics with fuzzy data. The statistical problem with random set observations is this.

Let X be a random vector, defined on a probability space (Ω, $\mathcal{A}$,$\mathcal{P}$), with value in R^m. The (unknown) probability law of X is the probability measure $\tau_0 = PX^{-1} = dF$ on $B(R^m)$, where F is the distribution function of X on R^m.

Let $X_1, X_2, ..., X_n$ be a (i.i.d.) random sample drawn from X. If this sample is observable then the Glivenko-Cantelli theorem asserts that τ_0 assumption M_o can be estimated by the empirical measure dF_n, where

$$F_n(x) = \frac{1}{n}\#\{i : X_i \le x\},\ \ x \in R^m$$

is the empirical distribution function based on $X_1, X_2, ..., X_n$. In other words for the sample size n sufficiently large, τ_0 will be in some neighborhood of dF_n.

The inference about τ_0 is delicate when instead of observing $X_1, X_2, ..., X_n$, we can only observe a random sample $S_1, S_2, ..., S_n$ of random sets drawn from a random set S (defined on (Ω, $\mathcal{A}$, P) with values in the space $\mathcal{F}(R^m)$ of closed subsets of R^m, such that $X_i \in S_i$, a.s., i=1,2,….

First, the distribution of S is characterized by its capacity functional T which can be estimated for n sufficiently large by the strong law of large numbers, by the sequence of empirical capacity functionals

$$T_n(K) = \frac{1}{n}\#\{i : S_i \cap K \ne \phi\}$$

Next, since X is an *a.s.* selector of *S*, we have

$$\tau_o \in Core(T) = \{\tau : \tau \le T\}$$

(set of probability measure τ on $B(R^m)$ such that $\tau(A) \le T(A)$, $\forall A \in B(R^m)$). Thus, inference about τ_0, based on $S_1, S_2, ..., S_n$ could be achieved if $\tau_o \in Core(T)$ a.s., for n sufficiently large. Clearly, this requires the a.s. convergence, in some sense of the sequence of sets $Core(T_n)$ to the $Core(T)$. This will imply that a sequence of estimators $\tau_n(S_1, S_2, ..., S_n)$ of τ_0, based on $Core(T_n)$, converges *a.s.* to τ_0 in the weak-star topology (see [98] for the concept of weak-star topology).

Let us elaborate the above statements. For complete details the reader is referred to [3] for the case where $X: \Omega \to V$ with $V \in R^m$ and finite, and to [24] for V being a compact, metric space. But first, an example showing that the case of a capacity functional is adequately modeled as the parameter space of the unknown distribution of an unobservable random vector.

Suppose the random variable of interest X is not directly observable, and instead we observe another random variable Y taking values in the compact subspace.

$$[0,a] \subseteq R, \text{ with } P(X \leq Y) = 1$$

Let F_X, F_Y denote the distribution functions of X, Y respectively. Then the stochastic ordering $X \leq Y$, a.s. , is equivalent to $F_Y(\cdot) \leq F_X(\cdot)$. Assume further that X and Y are independent, then

$$P(X \leq Y) = \int_0^a F_X(t) dF_{\underline{Y}}(t) = 1 \tag{5.1}$$

restricting the class of distributions dominating F_Y.

Under the above model, the unknown dF_X lies in the parameter space Θ, where

$$\Theta = \{dF : F(x) = 0 \text{ for } x < 0,\ F_V \leq F \text{ and } F \text{ satisfies (5.1)}\}$$

Now, let $S = [0, Y]$ be a closed random set on R^+. Then S is a coarsening of X, its capacity functional T (extended to Borel sets of [0,a] as $T(A) = \sup\{T(K) : K \text{ compact, } K \subseteq A\}$) is given by:

$$T(K) = P(S \cap K \neq \phi) = P(Y \geq \inf K) = 1 - F_Y(\inf K)$$

And, $Core(T) = \{dF : dF \leq T\} = \Theta$

Thus, in general, in the setting of estimation from coarse data, we focus on the core of the capacity functional of the random set defining the coarsening scheme. Of course, if the coarsening scheme consists of a random fuzzy set, then the whole machinery of set theory should be extended to fuzzy set theory.

Let U be a compact metric space, such as [0,1]. We denote by $\mathcal{M}(\upsilon)$ the space of all probability measures on U (i.e. probability measures defined on the Borel σ —field of U)

As stated earlier, the counter—part of dF_n is the $Core(T_n)$ which is a subset of $\mathcal{M}(\upsilon)$. Since $\tau_o \in Core(T)$, the estimation of τ_o, based on $S_1, S_2, \ldots\ldots, S_n$, will be based upon the approximation of $Core(T)$ by $Core(T_n)$, noting that $dF_n \in Core(T_n)$, a.s., for any n. We will indicate also that the rate of convergence of $Core(T_n)$ to $Core(T)$ is exponential. The foundations for large sample statistics with random sets will be presented in the next sections which should from the basics for fuzzy random sets by considering α —level sets.

5.4 Random Set Data on Finite Spaces

The statistical framework for inference with random set data on finite space was laid down in [3]. Specifically, let U be a finite set, say $U = \{u_1, u_2, \ldots, u_k\}$, i.e. $|U| = k$. Suppose the random variable X:(Ω, $\mathcal{A}$, $\mathcal{P}$) $\rightarrow U$ is unobservable, and instead, we observe a random set S:(Ω, $\mathcal{A}$, $\mathcal{P}$) $\rightarrow 2^U$ where X is an a.s. selector of S, i.e. $P(X \in S) = 1$. Let τ_0 denote the true, unknown probability law of X on U. In order to estimate τ_0 from an i.i.d. sample $S_1, S_2, \ldots\ldots, S_n$ from S, we need first to specify the parameter space Θ in which τ_0 lies.

Let T denote the capacity functional of S. Let Γ denote the class of all capacity functionals on 2^U. We regard Γ as a compact convex subset of $[0,1]^{2^k}$. Also, let $\mathcal{P}$ denote the class of all probability measure on U. We identify $\mathcal{P}$ with a compact, convex subset of $[0,1]^k$. Suppose τ_0 is known to belong to some "*model*" $\Phi \subseteq \mathcal{P}$, where Φ is assumed to be closed (therefore compact). The parameter space Θ will be $\Phi \cap Core(T)$. The explicit description of $Core(T)$ is useful for statistical inference about τ_0 as well as for the extension to the case of a compact, metric space U in Section 5.5.

While the fact that $Core(T)$ contains precisely all probability laws of the a.s. selectors of T, in a fairly general setting, say when U is a Hausdorff, locally compact, second countable topological space, e.g. $\mathcal{R}^d$, is a consequence of a general theorem of Noberg [12] on ordered coupling S of random sets, we will elaborate on this fundamental fact in the finite case.

Let S be a non-empty random set on a finite set U. Its probability density function is a map

$$f : 2^U \to [0,1] \text{ such that } f(\phi) = 0\,,\ \sum_{A \subseteq U} f(A) = 1$$

An allocation of f is a function

$$\alpha : U \times (2^U - \{\phi\}) \to [0,1]$$

such that $\forall A \subseteq U$, $\sum_{a \in A} \alpha(u, A) = f(A)$

For example, let $P : U \to [0,1]$ be a probability density (of some random variable taking value in U), such that $\forall u \in U$, $P(u) \neq 0$

Let

$$\alpha(u, A) = \frac{f(A)}{P(A)} P(u)$$

where

$$P(A) = \sum_{u \in A} P(u)$$

then $\alpha(\cdot,\cdot)$ is an allocation of f. In particular, if $P(u) = \dfrac{1}{|U|}$, let $\forall u \in U$, i.e. P is the uniform density on U, then

$$\alpha(u, A) = \frac{f(A)}{|A|}, A \neq \phi$$

Each allocation $\alpha(.,.)$ gives rise to a probability measure on U via the probability density function $g_\alpha : U \to [0,1]$ where

$$g_\alpha(u) = \sum_{u \in A} \alpha(u, A)$$

the summation is of course voer all $A \subseteq U$ which contain u.
The associated probability measure is

$$\tau_\alpha(A) = \sum_{u \in A} g_\alpha(u)$$

Note that, equivalently, we can characterize the probability law of a random set S, either by its probability density f on 2^U, or by its capacity functional T, where

$$T(A) = P(S \cap A \neq \phi)$$

or by its distribution function

$$F(A) = P(S \subseteq A) \text{ noting that } T(A) = 1 - F(A^C)$$

Clearing if X is an a.s. selector of a nonempty random set S, then

$$P_X \in Core(T)$$

The converse holds. In fact, the following result is sufficiently general.

Theorem 5.1 (Noberg, [12])

Let τ *be a probability measure on* $B(R^d)$ *and T be a capacity functional on* $R(R^d)$. *Then the following are equivalent:*

(i) $\tau \in Core(T)$

(ii) There exists a common probability space (Ω, $\mathcal{A}$,$\mathcal{P}$) *on which are defined a ramdom closed set* $S:\Omega \to F(R^d)$, *and a random variable* $X:\Omega \to R^d$ *such that* $P(X \in S)=1$, *and S and X have T and* τ *as their distributions, respectively.*

It turns out that elements of Core(T) all come from allocations. Specifically,

Theorem 5.2

Let T be a capacity functional on a finite set U. Let f be the Mobius inverse of the dual (distribution) F of T (i.e. f is the associated probability density on 2^U *). Then a probability measure* τ *on U is in Core(T) if and only if* τ *comes from an allocation of f.*

Proof

(a) Necessity. Suppose $\tau \in Core(T)$. By the above Noberg's theorem, there is a common probability space (Ω, $\mathcal{A}$,$\mathcal{P}$) on which are defined

$$S:\Omega \to 2^U \setminus \{\phi\}, \quad X:\Omega \to U$$

such that $P(S \in U)=1$, with S and X having T and τ as distributions, respectively.

Define

$$\alpha(u,A) = P(S=A, X=u)$$

Clearly, $\alpha(u,A)=0$ for $u \notin A$, and $\forall A \subseteq U$, $\sum_{u\in A}\alpha(u,A)=f(A)$ i.e. α is an allocation with

$$\tau(\{u\}) = P(X = u) = \sum_{u\in A}\alpha(u, A),$$

i.e. τ comes from the above allocation α

(b) Sufficiency. Consider the following probability space

$\Omega = U \times (2^U \setminus \{\phi\})$, $\mathcal{A}$=power set of Ω, $P(\{w\})= P((u,A)) = \alpha(u,A)$, where α is an allocation of *f*.

Define

$$S : \Omega \to 2^U \setminus \{\phi\}$$

$$X : \Omega \to U \text{ by } S(u,A) = A, \ \ X(u,A) = u, \text{ for any } (u,A) \in \Omega.$$

We have:

$$P(S = A) = P(A \times U) = \sum_{u\in U} P((u, A)) = f(A)$$

so that *f* is the density of *S* (i.e. T is the capacity functional of *S*).

On the other hand, the probability measure M_α arising from the allocation α is precisely the probability law of X. Indeed,

$$P(X = u) = P((\{u\} \times (2^U \setminus \{\phi\}))$$

$$= \sum_{\phi\neq A\subseteq U} P((u, A)) = \sum_{\phi\neq A\subseteq U}\alpha(u, A)$$

Moreover, *X* is an *P*-a.s. selector of *S*. Indeed,

$$P(X \in S) = P((u, A) : u \in A = \sum_{A\subseteq U}\sum_{u\in A}\alpha(u, A) = \sum_{A\subseteq U} f(A) = 1$$

Thus, $\tau_\alpha \in Core(T)$

Remarks

(i) As we will see, the above Theorem 5.2 for U finite is useful for the extension of statistical inference to more general cases.

(ii) Traditionally the above Theorem 5.2 can be proved by using Shapley's Theorem [83] in game theory. This can be seen as follows.

Let U with $|U| = k$, and f by the probability density function of a non-empty random set S.

Let $U = \{u_1, u_2, ...u_k\}$ be an ordering of U. Set $\alpha(u, A) = f(A)$ if $u = u_j \in A$ where $u_j = \max\{i : u_i \in A\}$, and zero otherwise.

Let $F(A) = \sum_{B \subseteq A} f(B)$ be the associated distribution function. Then the above special allocation α gives rise to the special probability density function g_α on U, namely:

$$g_\alpha(u_i) = \sum_{u_i \in A} \alpha(u, A) = \sum_{u_i \in A \subseteq \{u_1, u_2, ...u_i\}} f(A)$$

$$= F(\{u_1, u_2, ..., u_i\}) - F(\{u_1, u_2, ..., u_{i-1}\})$$

for i=1,2,…,k (if i=1,then $F(\{u_1, u_2, ..., u_{i-1}\}) = F(\phi) = 0$).

Let $\mathcal{P}$ denote the set of all probability measures on U. Since each probability measure $\tau \in \mathcal{P}$ is uniquely determined by its density function on U, we identify τ with a vector in the unit simple S_k of R^k where

$$S_k = \{p = (p_1, p_2, ..., p_k) \in [0,1]^k, \sum_{i=1}^{k} p_i = 1\}$$

By identification, *Core*(T) is a compact convex subset of S_K , where

$$T(A) = 1 - F(A^c)$$

There are *k!* different orderings of the elements of the set *U*. Let $\sum$ denote the set of all permutations of $\{1,2,\ldots,k\}$. For each $\sigma \in \sum$, the elements of *U* are indexed as $\{u_{\alpha(1)}, u_{\alpha(2)}, \ldots, u_{\alpha(k)}\}$, and $g_\sigma(u_{\sigma(1)}) = F(\{u_{\sigma(1)}\})$, and for $i \geq 2$,

$$g_\sigma(u_{\sigma(i)}) = F(\{u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(i)}\}) - F(\{u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(i-1)}\})$$

The associated probability measure τ (having g_σ as density on *U*) is in *Core*(*T*). Indeed, for simplicity, take $\sigma(i) = i$, for i=1,2,…,k. For $\tau(A) = \sum_{u \in A} g(u)$ (here we drop σ from our writing, but not from our mind !), we have $F(A) \leq \tau(A)$ when *A*=*U*.

For *A*≠*U*, let $j = \min\{i : u_i \in A^c\}$, noting that $A^c \neq \phi$. Let $B = \{u_1, u_2, \ldots, u_j\}$, then $A \cap B = \{u_1, u_2, \ldots, u_{j-1}\}$, $A \cup B = A \cup \{u_j\}$. Since F is monotone of order 2,

i.e. $F(A \cup B) \geq F(A) + F(B) - F(A \cap B)$, we have

$$F(A \cup \{u_j\}) \geq F(A) + F(\{u_1, u_2, \ldots, u_j\}) - F(\{u_1, u_2, \ldots, u_{j-1}\})$$

as $g(u_j) \leq F(A \cup \{u_j\}) - F(A)$.

But $g(u_j) = \tau(\{u_j\}) = \tau(A \cup \{u_j\}) - \tau(A)$

So that

$$F(A) - \tau(A) \leq F(A \cup \{u_j\}) - \tau(A \cup \{u_j\})$$

On the right hand side of the above inequality, viewing $A \cup \{u_j\}$ as another set *A'*, and using the same reasoning, we have

$$F(A \cup \{u_j\}) - \tau(A\{u_{j\}}) = F(A') - \tau(A') \leq F(A \cup \{u_i\}) - \tau(A \cup \{u_i\})$$

where $u_i \neq u_j$. Continuing this process, we arrive at

$$F(A) - \tau(A) \le F(A \cup \{u_j\}) - \tau(A \cup \{u_j\})$$

$$\le F(A \cup \{u_j, u_i\}) - \tau(A \cup \{u_j, u_i\})$$

$$\le ... \le F(U) - \tau(U) = 1 - 1 = 0$$

Thus, $F \le \tau$, i.e. $\tau \in Core(T)$.

Now, for each $\sigma \in \Sigma$, $\tau_\sigma \in Core(T)$. There are *k!* (not necessarily distinct) τ_σ .

Shapley Theorem 5.3 [43]. *Let T be the Capacity functional of some non-empty random set as U with* $|U| = k$ *. Then Core(T) is a compact, convex polyhedron with at most k! extreme points which are precisely the* τ_σ*'s* .

The above results imply that Core(T) consists precisely of probability measures coming from allocations of *f* (the *Mobius* inverse of $F(A) = 1 - T(A^c)$. Indeed , let $\mathcal{A}$ denote the subset of $\mathcal{P}$ consisting of all probability measures coming from *f*. Let α be an allocation of *f*. Then, for any $A \subseteq U$

$$F(A) = \sum_{B \subseteq A} f(B) = \sum_{B \subseteq A} \sum_{u \in B} \alpha(u, B)$$

$$\le \sum_{u \in A} \sum_{u \in B} \alpha(u, B) = \tau_\alpha(A)$$

i.e. $\mathcal{A} \subseteq Core(T)$

Conversely, by Shapley's theorem, *Core*(*T*) has the τ_σ*'s* as extreme points and hence *Core*(*T*) is the set of convex combinations of these extreme points. Since P_σ's are elements of the convex set $\mathcal{A}$, $\mathcal{A}$ contains all convex combinations of its elements, in particular, convex combinations of these τ_σ*'s* which is *Core*(*T*).

We are now ready to present Schreiber's work [81]

With the previous notation, the essence of statistical inference about τ_0 is to establish a framework in which we can assert that, based on the observations of $S_1, S_2, ..., S_n$, estimates of τ_0 can be constructed in a consistent manner. For this, first we need to define concepts of "*distance*". Recall that $\mathcal{P}$ denotes the class of all probability measures on the finite set *U*. A discrepancy measure on $\mathcal{P}$ is a map

$$\Delta : \mathcal{P} \times \mathcal{P} \to R^+ \text{ such that}$$

(1) $\Delta(\tau, \upsilon)=0 \Leftrightarrow \tau = \upsilon$

(2) Δ is lower semi-continuous ($\mathcal{P}$ is identified with compact convex of $[0,1]^k$

(3) Δ is strictly convex on $\mathcal{P} \times \mathcal{P}$

(4) for each $\upsilon \in \mathcal{P}$, the map $\tau \in \mathcal{P} \to \Delta(\tau, \upsilon)$ is continuous.

For example, the following are discrepancy measures：

(a) $\Delta^\infty(\tau, \upsilon)=\max\{|\tau(u_i) - \upsilon(u_i)| : i = 1,2,...,k\}$ (i.e. the l^∞-distance).

(b) For $p \geq 1$

$$\Delta^p(\tau, \upsilon)=[\sum_{i=1}^{k}|\tau(u_i) - \upsilon(u_i)|^p]^{\frac{1}{p}}]$$

Let Δ be a given discrepancy measure on $\mathcal{P}$. From axioms (2) and (3), it follows that, for each $\upsilon \in \mathcal{P}$., there is a unique probability measure in $\mathcal{P}$, denoted as $\upsilon(T)$ such that

$$\Delta(\upsilon(T), \upsilon) = \inf\{\Delta(\tau, \upsilon) : \tau \in Core(T)$$

Thus, let us extend Δ to $\Gamma \times \mathcal{P}$ (recall that Γ is the class of all capacity functional on U) as∶

$$\Delta^*(T,\upsilon) = \inf\{\tau(\tau,\upsilon) : \tau \in Core(T)\}$$

where $\Delta^*(T,\upsilon) = \Delta(\tau(T),\upsilon)$, $\forall\ \upsilon \in \mathcal{P}$. This means that $\upsilon(T)$ is the probability law of the a.s. selector (of the random set *S* having *T* as its capacity functional) which is the best approximation of *U* in the Δ sense. In particular, if V is itself a probability law of some a.s. selector of *S*, thus by (1), $\Delta^*(T,\upsilon) = 0$ and $\upsilon(T) = \upsilon$. The extension Δ^* has nice topological properties.

Specifically, $\Delta^* : \Gamma \times\ \mathcal{P} \to R^+$ is lower semi-continuous and strictly convex. Furthermore for each fixed $\upsilon \in \mathcal{P}$, the map $T \in \Gamma \to \Delta^*(T,\upsilon)$ is continuous. (See [3] for details).

Now, as in [3], let Φ be a statistical model for τ_0. Define for each $T \in \Gamma$,

$$\hat{\Delta}(T \mid \Phi) = \inf\{\Delta^*(T,V) : V \in \Phi\}$$

$$\hat{\Delta}(T \mid \Phi) = \inf\{\Delta^*(\tau,\upsilon) : \upsilon \in \Phi\}$$

$$= \inf\{\Delta(\tau,\upsilon) : \tau \in Core(T), \upsilon \in \Phi\}$$

It follows from topological properties of Δ^* and the compactness of Φ that there is (not necessarily unique) an $\upsilon \in \Phi$ such that

$$\Delta^*(T,V) = \hat{\Delta}(T \mid \Phi)$$

For a such υ, we have $\Delta(\upsilon(T),\upsilon) = \hat{\Delta}(T \mid \Phi)$, and hence υ (*T*) is viewed as the most likely probability law governing the unobserved data leading the deserved random set data.

Thus, consider $\Phi^* \subseteq \Phi$, where $\Phi^* = \{\upsilon \in \Phi : \Delta^*(T,\upsilon) = \hat{\Delta}(T \mid \Phi)\}$. Note that Φ^* is compact. The interpretation of Φ^* is that *A* is the class of probability laws on V which fit best the observations $S_1, S_2, \ldots, S_n$.

Thus, it boils down to identity the elements of the “optimal” set Φ^*.

Recall that the empirical capacity functional $T^{(n)}$, based on $S_1, S_2, \ldots\ldots, S_n$ is.

$$T^{(n)}(A) = \frac{1}{n}\sum_{i=1}^{n}(S_i \cap A \neq \phi)$$

Let $\Phi_n^* = \{\upsilon \in \Phi : \Delta^*(T^{(n)}, \upsilon) = \hat{\Delta}(T^{(n)} \mid \Phi)\}$. It can be checked that Φ_n^* is a.s. non-empty and compact. For $\phi \neq A \subseteq P$, and $\upsilon \in P$, define

$$d(\upsilon, A) = \inf\{\Delta^{(2)}(\upsilon, \tau) : \tau \in A\}$$

On the metric space (P, $\Delta^{(2)}$) we have $d(\upsilon, A) = 0 \Leftrightarrow \upsilon \in A$, and the map $\upsilon \to d(\upsilon, A)$, for each $A \neq \phi$, A compact in P, is continuous. The following facts are essential.

Let $(T_n, n \geq 1) \subseteq \Gamma$ such that $T_n \to T \in \Gamma$, and

$$\Phi^{(n)} = \{\upsilon \in \Phi : \Delta^*(T_n, \upsilon) = \hat{\Delta}(T_n \mid \Phi)\} \text{ for } n \geq 1$$

Then, with probability one, we have:

(i) $\Phi^{(n)} \subseteq \Phi$ is non-empty and compact

(ii) $\lim_{n\to\infty} \hat{\Delta}(T_n \mid \Phi) = \hat{\Delta}(T \mid \Phi)$

(iii) $\lim_{n\to\infty} \sup\{d(\upsilon, \Phi) : \upsilon \in \Phi^{(n)}\} = 0$

These facts imply Schreiber's basic result:

Theorem 5.4 (Schreiber [3]) *With probability one, we have*

$$\lim_{n\to\infty} \hat{\Delta}(T^{(n)} \mid \Phi) = \hat{\Delta}(T \mid \Phi), \quad \lim_{n\to\infty} \sup\{d(\upsilon, \Phi) : \upsilon \in \Phi^{(n)}\} = 0,$$

where $\{T^{(n)}, n \geq 1)$ *is the sequence o empirical capacity functionals based on* $S_1, S_2, \ldots, S_n$.

In addition, the rate of convergence of $\hat{\Delta}(T^{(n)} \mid \Phi)$ to $\hat{\Delta}(T \mid \Phi)$ can be specified. Specifically, for each ε>0, there is an $L_\varepsilon > 0$ such that, for n sufficiently large,

$$P(|\hat{\Delta}(T^{(n)} \mid \Phi) - \hat{\Delta}(T \mid \Phi)| \geq \varepsilon) \leq e^{-nL_\varepsilon}$$

Let us elaborate a little bit on this result of Schreiber which is based on large deviations principle (see e.g. [15]).

Let $X_1, X_2, ..., X_n$ be an *i.i.d.* sample form the Bernouilli random variable X taking value in $\{0,1\}$ with *P*(*X*=1)=*p*. The sample mean $\overline{X}_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$ is a weakly consistent estimator of *EX* = *p*, since $\forall \varepsilon > 0$,

$$P(|\overline{X}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \to 0 \text{ as } n \to \infty.$$

The rate of which $\overline{X}_n \to p$ can be made explicit as follows. The events $\{\left|\overline{X}_n - p\right| \geq \varepsilon\}$ are very rare events in the sense that then probabilities decay to zero exponentially fast. They represent large deviations of $\overline{X}_n$ from the mean *EX* =*P* since $\overline{X}_n$ avoids some neighborhood of *EX*, whereas the event $\{|\overline{X}_n - p| < \varepsilon\}$ represent small deviations of $\overline{X}_n$ from *P*, i.e. $\overline{X}_n$ is in some neighborhood of *P*.

Let *F* be the distribution function of *X*, and $\varphi(t) = E(e^{tx}) = \int_R e^{tx} dF(x)$, $t \in R$ be the moment generating function of *X* (or the Laplace transform of *F*). Here, for the Bernoulli case, $\varphi(t) = 1 - p + pe^t$.

Let $\Lambda(x) = \sup\{xt - \log\varphi(t) : t \in R\}, x \in R$, be the so-called Cramer transform of *F*. The function Λ is called a rate function. The following are well-known. Let P_n denote the probability law of $\overline{X}_n$. Then：

(a) For every closed set *A* of *R*,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(A) \leq -\inf\{\Lambda(x) : x \in A\}.$$

(b) For every open set B of R,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(B) \geq -\inf\{\Lambda(x) : x \in B\}.$$

Let $I(A) = \inf\{\Lambda(x) : x \in A\}$. Then, $\forall A \in B(R)$, $P_n(A)$ decays exponentially fast with rate $I(A)$, i.e.

$$\lim_{n\to\infty} \frac{1}{n} \log P_n(A) = -I(A).$$

On finite dimension spaces like R^d, the exponential rate of convergence is typically given by the Cramer transform which is defined in terms of the Lebesgue measure (as a reference measure on $\mathcal{B}B(R^d)$). Large deviations of sample paths of stochastic processes such as Brownian motions, require infinite dimensional spaces, like C[0,1], the space of continuous functions defined on the unit interval [0,1].

In order to cover all cases, a general large deviations principle is formulated without referring to any reference measure on the space. To this end, the Cramer transform is abstracted from its basic properties. The general large deviations principle is as follows principle is as follows.

Let U be a polish (i.e. a complete, separable, metric space) and Ψ its Borel σ-field. A rate function is a function $I : U \to [0,\infty)$, which is lower semi-continuous. A family ($P_\varepsilon, \varepsilon > 0$) of probability measures on Ψ is said to satisfy the large deviations principle with rate function I if for any $A \in \Psi$,

$$-\inf_{\overset{o}{A}} I \leq \liminf_{\varepsilon\to 0} \varepsilon \log P_\varepsilon(A)$$

$$\leq \limsup_{\varepsilon\to 0} \varepsilon \log P_\varepsilon(A) \leq -\inf_{\overline{A}} I$$

where $A^o, \overline{A}$ denote the interior and closure of A, respectively. This is equivalent to

$$\limsup_{\varepsilon\to 0} \varepsilon \log P_{\varepsilon}(A) \leq -\inf_{\bar{A}} I\,,\ \forall A \text{ closed, and}$$

$$\limsup_{\varepsilon\to 0} \varepsilon \log P_{\varepsilon}(A) \geq -\inf_{\bar{A}} I\,,\ \forall A \text{ open,}$$

where $I(A) = \inf\{I(x) : x \in A\}$.

For example, Cramer large deviations principle on R is extended to R^d where the Fenchel-Legender transform takes the role of rate functions. Specially, let F be a distribution function on R^d. The Laplace transform of dF is

$$\varphi(t) = Ee^{<t,x>}$$

$$= \int_{R^d} \exp[\sum_{j=1}^{d} t_j x_j] dF(x)$$

where $t = (t_1, t_2, ..., t_d) \in R^d$, $x = (x_1, x_2, ... x_d) \in R^d$. The Fenchel-Lebesgue transform of dF is

$$\Lambda(t) = Sup\{< t, x > - \log \varphi(t) : x \in R^d\}.$$

For example, let S be a non-empty random set on a finite set U with capacity functional T. Since $T(A) = P(S \cap A \neq \phi)$, we can consider the associate random vector X with values in R^d with $|U| = d$, namely

$$X = \begin{pmatrix} 1_{(S\cap A_1 \neq \phi)} \\ \vdots \\ 1_{(S\cap A_d \neq \phi)} \end{pmatrix}$$

where A_j's, j = 1,2,…,d, are all subsets of U. If we identify each map $h : 2^U \to R$ with a vector in R^d, namely

$$\begin{pmatrix} h(A_1) \\ \vdots \\ h(A_d) \end{pmatrix} = \begin{pmatrix} t_1 \\ \vdots \\ t_d \end{pmatrix} = t,$$

then we can consider the Laplace transform of X as a function defined on maps h:

$$\varphi_S(h) = \varphi_X(t) = Ee^{<t,x>}$$

$$= E\exp[\sum_{j=1}^{d} h(A_j) 1_{(S \cap A_j \neq \phi)}]$$

$$= E\exp[\sum_{S \cap A_j \neq \phi} h(A)]$$

Also by identifying the capacity functionals T with the vector $(T(A), A \subseteq U)$ in R^d, the Fenchel –Legendre transform of X is written as:

$$\Lambda_S(T) = Sup\{\sum_{A \subseteq U} h(A)T(A) - \log \varphi_S(h) : h : 2^U \to R^d\}$$

with these identifications, Schreiber's result is stated as follows:

The sequence $\{\hat{\Delta}(T^{(n)} \mid \Phi), n \geq 1\}$ satisfies the large deviations principle with rate function given by $h \to \inf\{\Lambda_S(T) : T \in \Gamma, \hat{\Delta}(T \mid \Phi) = h\}$.

5.5 Random Set Data on Metric, Compact Spaces

As stated earlier statistics with set-values observations are not only new but also essential in developing framework for statistics with fuzzy set observations, via, say techniques based on α-level sets of fuzzy sets. As such, it is useful for researchers who wish to investigate statistical methods for fuzzy data to be aware of current developments in framework for statistical inference with random set data. This section is devoted to the extension of finite random set models in 5.4 to the more general case of metric, compact spaces. The material presented here is drawn from the recent work [23].

Let (U, d) be a metric, compact space, such as [0,1], and Ψ its Borel σ-field. All random (closed) sets and variables are assumed to be defined on a probability space (Ω, $\mathcal{A}$, $\mathcal{P}$). As in section 5.4, we look at the situation of coarse data in which we observe a random sample $S_1, S_2, ..., S_n$ from a non-empty random set model S which is a random closed (in fact here compact) set, i.e. taking values in the class of closed subsets $\mathcal{F}$ (=$\mathcal{K}$, the class of compact subsets) of U. The unknown capacity functional of S is denoted as T. Let $\mathcal{P}$ denote the class of all probability measures on Ψ. Since the core of T, i.e.

$$Core(T) = \{\tau \in \mathcal{P},: \tau \leq T\}$$

is non-empty, there exists, according to Noberg's theorem (section 5.4), a random variable $X : \Omega \to U$ which is an a.s. selector of S, i.e $P(X \in S) = 1$.

We can view X as the random variable of interest but unobservable. The most general statistical problem is the inference about the unknown probability law τ_0 of X (a probability measure on Ψ) from the observations $S_1, S_2, ..., S_n$. We follow the set–up in Section 5.4.

The estimation of τ_0 is essentially based upon the convergence of the sequence of cores of empirical functionals $T^{(n)}$ to *Core* (T) which is a compact subset of $\mathcal{P}$, where $\mathcal{P}$ is a topological space with the weak-star topology (see e.g. [41]). Specially, a metric on $\mathcal{P}$ is defined as:

$$\Delta(\tau, \upsilon) = \sum_{i=1}^{\infty} \frac{| Sf_n d\tau - \int f_n d\upsilon |}{2^n \, \| f \|}$$

where $\{f_n, n \geq 1\}$ is a sequence of continuous, real-valued functions dense in the space $C(U)$ of continuous, real-valued functions defined on U (endowed with the uniform topology), and

$$\| f \| = \max\{| f(x) | : x \in U\}$$

In fact, $\{\mathcal{P}, \Delta\}$ is a compact space. To generalize Schreiber's result in the finite case, we proceed as follows. We define a pseudo-metric on the space Γ of all capacity functionals on $\mathcal{K}$, and show that $Core(T)$ depends continuously on T. These mathematical results lead to the existence of strong consistent estimators $(\tau_n, n \geq 1)$ of τ_0 in the weak star convergence sense. Specially, let $\tau_0 \in Core(T)$ and $T^{(n)}$ be the empirical capacity functionals based on $S_1, S_2, ..., S_n$. Then a.s., there exists $\tau_n \in Core\ (T^{(n)})$, $n \geq 1$, such that $\tau_n \to \tau_0$ in the weak star convergence sense as $n \to \infty$.

Moreover, let e_Δ denote the Hausdorff metric on non-empty, compact subsets of $\mathcal{P}$, i.e.

$$e_\Delta(A,B) = \inf\{\varepsilon \geq 0 : N_\varepsilon(B) \supseteq A, N_\varepsilon(A) \supseteq B\}$$

where $N_\varepsilon(A)$ is the neighborhood of A in $\mathcal{P}$. i.e.

$$N_\varepsilon(A) = \{\tau \in \mathcal{P} : \exists \upsilon \in A, \text{ such that } \Delta(\tau, \upsilon) < \varepsilon\}$$

Then for any $\varepsilon > 0$, there exist n_ε and $L_e > 0$ such that

$$P(e_\Delta(Core(T), Core(T^{(n)})) > \varepsilon) \leq e^{-nL_\varepsilon}$$

for all $n \geq n_\varepsilon$.

Here is the sketch of the proof.

On the metric space (U,d), a set $A \subseteq U$ is called a ε-spanning set of U if $N_\varepsilon(A) = U$, for $\varepsilon > 0$. Since U is compact, for each $n \geq 1$, there is a finite *1/n*-spanning set of U, denoted as H_n.

Let $\theta_n = \{N_{1/2}(A) : A \subseteq H_n\}$, $n \geq 1$. Then a pseudo-metric on Γ (space of all capacity functionals on $\mathcal{K}(U)$ is defined as:

$$\Delta^*(T,T') = \sum_{i=1}^{\infty} 2^{-n-|H_n|} \sum_{A \in Q_n} |T(A) - T'(A)| \ .$$

It can be proved that if $T_n \to T$ in Δ^* – sense, then $Core(T) \to Core(T)$ in e_Δ-sense. As an application, if $T_n = T^{(n)}$, i.e. T_n is the sequence of empirical capacity functionals, then for any $\tau_0(w) \to \tau$ is Δ - sense.

Remark: Using the characterization of the core of a capacity functional on a finite set (in terms of allocations) and the “finite approximation” of capacity functionals on compact, metric spaces by capacity functionals on finite set, the above estimators $\tau_n(w)$ are in fact constructive. For details, see [4].

As in the finite case, the exponential rate of convergence of $Core(T^{(n)})$ to $Core(T)$ in e_Δ-sense, is obtained by using classical large deviations principle for i.i.d random vectors.

Chapter 6 Convergence of Random Fuzzy Sets

As in Chapter 5, we view random fuzzy sets as generalizations of random closed sets on $\mathcal{R}^d$, or more generally on Hausdorff, locally compact, second countable topological spaces. Thus, technically speaking, the study of limit theorems for random fuzzy sets can be carried out by using α-level sets and results from limit theorems for random closed sets. For this reason, we focus in this chapter on convergence of random closed sets.

6.1 Stochastic Convergence of Random Sets

Again let $(\Omega, \mathcal{A}, \mathcal{P})$ be a probability space, and $\mathcal{F}(\mathcal{R}^d)$ or simply $\mathcal{F}$, be the set of closed subsets of $\mathcal{R}^d$. Let τ denote the hit-or miss topology on $\mathcal{F}$, and $\mathcal{B}(\mathcal{F})$ its associated $Borel\, \sigma - field$. Equipped with τ, $\mathcal{F}$ is a compact, metrizable space.

A random closed set is a map from Ω to $\mathcal{F}$ which is $\mathcal{A}$-$\mathcal{B}(\mathcal{F})$ measurable. Let $S_1, S_2, ..., S_1,$ be an i.i.d sample from a random closed set S. Large sample statistics relies on limit behavior of functions of the $S_i^{'}s$. For example, the laws of large numbers deal with the limits, in probability and almost surely, of "*sample means*" $(S_1 + S_2 + ... + S_n)/n$ in the sense of Minkowski.

Recall that, the Minkoski's operations for subsets of $\mathcal{R}^d$, on more generally of a linear space, are defined as follows:

For A, $B \subseteq \mathcal{R}^d$, and $\alpha \in \mathcal{R}$,

$$A + B = \{x + y : x \in A, y \in B\}$$

$$\alpha A = \{\alpha x : x \in A\}$$

Let $Z_n = (S_1 + S_2 + ... + S_n)/n$. We are interested in formulating stochastic convergence concepts for set-valued random elements $(Z_n, n \geq 1)$.

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 71–108 (2006)
www.springerlink.com

For simplicity, we restrict ourselves to sequences of random sets (convergence of sets can be formulated similarly)

Let S, S_n be random closed sets on $\mathcal{R}^d$. Then by $S_n \xrightarrow{a.s} S$, as $n \to \infty$, we mean

$$P(w : S_n(w) \to S(w)) = 1, \tag{6.1}$$

where $S_n(w) \to S(w)$ in the sense of the topology τ.

Note that the convergence in the topology τ is the same as the Painleve-Kuratowski convergence (see e.g. Salinetti and Wets [76], Molchanov [54]), namely

$$S(w) = \lim_n \sup S_n(w) = \lim_n \inf S_n(w) \tag{6.2}$$

where

$$\lim_n \sup S_n(w) = \{ x \in R^d : x_{n(k)} \to x : \text{for } x_{n(k)} \in S_{n(k)}, n(k), k \geq 1 \text{ being a subsequence}\} \tag{6.3}$$

$$\lim_n \inf S_n(w) = \{ x : x_n \in S_n(w), n \geq 1, x_n \to x \}. \tag{6.4}$$

Note that set-theoretic concepts of limits of sequences of sets are different than the above concepts of *limsup* and *liminf*. See also Klein and Thompson [40].

Remark: In the context of Banach spaces, random fuzzy sets can be embedded into Banach spaces, where the concept of Mosco convergence can also be need. See Li et al [45].

Since $\mathcal{R}^d$ is a metric space, with metric denoted as ρ, the concept of convergence in probability of sequences of random closed sets can be formulated

For $\varepsilon > 0$, and $A \subseteq \mathcal{R}^d$, $x \in \mathcal{R}^d$, $\rho(x, A) = \inf\{\rho(x, y) : y \in A\}$ and $A^{\varepsilon} = \{x \in R^d : \rho(x, A) < \varepsilon\}$

A sequence S_n of random closed sets on $\mathcal{R}^d$ is said to converge in probability to the random closed set S if for every $\varepsilon > 0$, and compact set $K(\mathcal{R}^d)$,

$$\lim_{n\to\infty} P\{w : ((S_n(w) \setminus S^{\varepsilon}(w)) \bigcup (S(w) \setminus S_n^{\varepsilon}(w)) \bigcap K \neq \phi\} = 0.$$

For a flavor of research on laws of large numbers for random sets and random fuzzy sets, see e.g. Artstein and Vitale [46] and Dozzi et al. [48]. A central limit theorem for random sets was first investigated by Cressie [14]. We discuss in same details next the convergence in distribution of random closed sets.

6.2 The Choquet Integral

The convergence in distribution of S_n to S is the usual weak convergence of probability measures on metric spaces. Specially, let PS^{-1} and PS_n^{-1} be probability laws of S and S_n, respectively, on $\mathcal{B}(\mathcal{F})$. Then $S_n \xrightarrow{D} S$, when $PS_n^{-1} \to PS^{-1}$ in the weak topology of the space of probability measures on $(\mathcal{F},\mathcal{B}(\mathcal{F}))$, see e.g. Billingsley [5]. While there are equivalent ways to study the convergence in distribution for random closed sets, we choose to formulate this type of convergence as a natural extension of the process based on Lebesque-Stietltje's theorem for random vectors, via the Chaquet theorem for random closed sets.

This program requires the concept of *Choque integral* which we elaborate in this section. First. Note that, as we will see, not only the concept of Choquet integral is useful for defining the convergence in distribution of random sets but it provides also an interesting tool for general decision–making.

Let X be a non-negative random variables, defined on $(\Omega, \mathcal{A}, \mathcal{P})$ Then its expected value can be written as:

$$E_P(x) = \int_{\Omega} X(w)dP(w)$$

$$= \int_{R^+} xdF(x)$$

$$= \int_0^\infty P(w : X(w) > t)dt \tag{6.5}$$

For $X : \Omega \to \mathcal{R}$, we can write

$$E_P(X) = E_p(X^+) - E_p(X^-)$$

$$\int_0^\infty P(X > t)dt + \int_{-\infty}^0 [P(X > t) - 1]dt, \tag{6.6}$$

where F is the distribution function of X, and X^+, X^- are positive and negative parts of X, respectively.

If we look at the term $\int_0^\infty P(X > t)dt$, we recognize that the function $t \to P(X > t)$ is measurable since it is a non-increasing function on R^+, and this last fact is due to the monotonicity of the set function P. Thus if P is replaced by any set function τ on $\mathcal{A}$ which is monotone, i.e. $A, B \in \mathcal{A}$. $A \subseteq B \Rightarrow \tau(A) \le \tau(B)$, then formally the integral

$$\int_0^\infty \tau(w : X(w) > t)dt$$

is well-define.

Moreover if $\tau(\Omega) = 1$, then the quantity:

$$C_\tau(X) = \int_0^\infty \tau(x > t)dt + \int_{-\infty}^0 [\tau(x > t) - 1]dt \tag{6.7}$$

is also well-defined.

$C_\tau(X)$ is termed the *Choquet integral* of the measurable function $X : \Omega \to R$ with respect to the monotone set function τ. This concept of "*integral*", extending Lebesgue integral with respect to σ-additive set functions (probability measures), is due to Choquet [12]. While this type of integral is widely used in mathematical economics, see e.g. Marinacci and

Montruccchio [50], it remains somewhat unfamiliar to statistics, see also e.g. Huber [38], Huber and Strassen [37].

As we will see, when replacing distribution functions F of random variables by capacity functionals T of random closed sets (via Choquet theorem characterizing of probability laws of random closed set in terms of capacity functionals which are not σ- additive set functions) Choquet integral with respect to capacity functionals is necessary.

Here is a situation in coarse data analysis. Suppose that for each random experiment w, we cannot observe the exact outcome $X(w)$, but we can only locate $X(w)$ in some closed subset of $\mathcal{R}$. Let S be *a coarsening of* X. Let g: $R \to R^+$, measurable. Suppose we are interested in computing $Eg(X)$ in this situation. The random variable $g(X)$ is bounded by

$$g_*(w) = \inf\{g(x) : x \in S(w)\} \le g(X(w))$$

$$\le \sup\{g(x) : x \in S(w)\} = g^*(w) \tag{6.8}$$

For g_* and g^* should be such that, for any $B \in \mathcal{B}(\mathcal{R})$,

$$B_* = \{w : S(B) \subseteq B\} \in \mathcal{A}$$

and

$$B^* = \{w : S(B) \bigcap B \neq \phi\} \in \mathsf{A}$$

In such a case, let

$F_* = B(R) \to [0,1]$ be defined as

$$F_*(B) = P\{w : S(w) \subseteq B\} = P(B_*) \tag{6.9}$$

Then

$$E_*g(X) = Eg_* = \int_\Omega g_*(w)dP(w)$$

$$= \int_0^\infty P\{w : g_*(w) > t\}dt$$

$$= \int_0^\infty P(g_*^{-1}(t,\infty))dt$$

$$= \int_0^\infty P(g^{-1}(t,\infty))_* dt$$

$$= \int_0^\infty P\{w : S(w) \subseteq g^{-1}(t,\infty) > t\}dt$$

$$= \int_0^\infty F_*(g^{-1}(t,\infty))dt$$

$$= \int_0^\infty F_*(x : g(x) > t)dt \tag{6.10}$$

Similarly, for

$$F^*(B) = P(S \cap B \neq \phi) = P(B^*), \tag{6.11}$$

we have

$$E^* g(X) = Eg^* = \int_0^\infty F^*(x : g(x) > t)dt\ . \tag{6.12}$$

Thus, both E_{g*} and E_g^* are Choquet integrals of the measurable function g with respect to monotone set functions F_* , F^* , respectively, and $E^* g \leq Eg(X) \leq Eg^*$.

The above situation with imprecise information is considered in Dempester [16].

Here is another simple problem in decision making with incomplete information.

Let X be a random variable with values in $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\} \subseteq R^+$, with probability density function f_0 partially specified as $f_0(\theta_1) \geq 0.4$, $f_0(\theta_2) \geq 0.2$, $f_0(\theta_3) \geq 0.1$.

To make decisions based upon expected value of X, we can only rely on $\inf\{E_f(X) : f \in P\}$, where $E_f(X)$ denotes the expected value of X with respect to the density f , and $\mathcal{P}$ denotes the class of all probability density functions on Θ satisfying the above partial specification.

Renaming the element of Θ if necessary, suppose $\theta_1 < \theta_2 < \theta_3 < \theta_4$. Let

$$F : 2^{\Theta} \to [0,1]$$

$$F(A) = \inf\{P_f(A) : f \in P\}, \tag{6.13}$$

where $P_f(A) = \sum_{\theta \in A} f(\theta)$.

Define $g : \Theta \to [0,1]$ by

$$g(\theta_1) = F(\{\theta_1, \theta_2, \theta_3, \theta_4\}) - F(\{\theta_2, \theta_3, \theta_4\})$$

$$g(\theta_2) = F(\{\theta_2, \theta_3, \theta_4\}) - F(\{\theta_3, \theta_4\})$$

$$g(\theta_3) = F(\{\theta_3, \theta_4\}) - F(\{\theta_4\})$$

$$g(\theta_4) = F(\{\theta_4\})$$

Clearly $g \in \mathcal{P}$. Now,

$$E_g(X) = \int_{\Theta} X(\theta) dP_g(\theta)$$

$$= \int_0^{\infty} P_g(\{\theta_i : \theta_i > t\}) dt = \int_0^{2} F(\{\theta_i : \theta_i > t\}) dt, \tag{6.14}$$

i.e. $E_g(X)$ is a Choquet integral.

We are going to show that $inf\{E_f(X) : f \in P\}$ is attained precisely at g. Since

$$E_f(X) = \int_0^{2} P_f(\{\theta_i : \theta_i > t\}) dt, \tag{6.15}$$

it suffices to show that, for each $t \in \mathcal{R}$,

$$E_f(X) = \int_0^2 P_f(\{\theta_i : \theta_i > t\})dt \text{, for all } f \in P \ . \tag{6.16}$$

If $\{\theta : \theta > t\}$ $\{\theta_1 < \theta_2 < \theta_3 < \theta_4\}$, then

$$P_g(\theta > t) = \sum_{i=1}^{4} g(\theta_i) = 1 = P_f(\theta > t) \ . \tag{6.17}$$

If $\{\theta : \theta > t\} = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, then

$$P_g(\theta > t) = \sum_{i=1}^{4} g(\theta_i) \le \sum_{i=2}^{4} f(\theta_i) = P_f(\theta > t) \ , \tag{6.18}$$

Since, by construction of g, we have

$$g(\theta_i) \le f(\theta_i) \text{ , for } f \in P \text{ , and } i = 2, 3, 4 \ . \tag{6.19}$$

Thus, the Choquet integral of $X(\theta) = \theta$ with respect to the monotone set function F is the lower bound for expected values in the model.

It is interesting to note the concept of Choquet integral is popular in the field of *multicriteria decision-making* (see e.g. Grabisch et al. [30]). Specifically, ranking of alternatives with respect to a set of criteria is a common and important process in almost every day activities. When the criteria are non-interactive (independent), then weighted a averages seems to be appropriate aggregation operators for ranking. Even in more complex situations, such as those involving fuzzy relations, linear (additive) aggregation operators are still in use, such as the popular *Analytic Hierarchy Process* (see Saaty [75]).

Recently, non-linear aggregation operators for ranking alternatives are shown to be more appropriate for interactive criteria, and the Choquet integral seems to be the most useful one, as it is a generalization of linear aggregation operators based on probability measures. In this context, the finite version of Choquet integral is used.

Specifically, let $C=\{1,2,...,k\}$ be a set of criteria. Let $X: C \to \mathcal{R}$ be an "evaluation". The degrees of importance of subsets of C is modeled by a set function $A: 2^C \to [0,1]$ such that $A(\phi)$=0, A(C)=1, and A is monotone, i.e. $A \subseteq B \Rightarrow \tau(A) \le \tau(B)$. Such a general set function is called a *fuzzy measure*.

Then,

$$C_\tau(X) = \sum_{i=1}^{k} x_{(i)} \, [\, \tau\,(A_{(i)}) - \tau\,(A_{(i+1)})\,], \tag{6.20}$$

where $x(i) = x_i$, $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$, $A_{(i)} = \{(i),(i+1),...,(k)\}$, $A_{(k+1)} = \phi$.

The discrete version $C_\tau(X)$ of the Choquet integral of X with respect to τ is used as a map from $R^n \to R$ to compare various evaluations: the evaluation X is preferred to the evaluation Y if $C_\tau(Y) \le C_\tau(X)$.

In the next sections, we are going to use Choquet integral to investigate the convergence in distribution of random sets. Statisticians who wish to know more about the mathematics of Choquet integral could see e.g. Grabisch et al [30], Nguyen and Walker [62], Marinacci and Martrucchio [50].

6.3 A Variational Calculus of Set Functions

While coarse data analysis seems to be an important and general view that coarsening schemes are random sets (or more generally random fuzzy sets) has not been sufficiently emphasized. Moreover, distributional aspects of large sample statistics based on random sets, let alone random fuzzy sets, have not yet entered statistical analysis. A typical situation is the estimation of probability density function using the excess mass approach (see e.g. Polonik [28]).

Recall that, like membership functions of fuzzy sets, a probability density function $f: R^d \to R^+$ can be recovered from its $\alpha - level$ sets

$$A_\alpha(f) = \{x \in R^d : f(x) \geq \alpha\}, \alpha \in R^+ \tag{6.21}$$

via

$$f(x) = \int_0^\infty A_\alpha(f)(x)d\alpha, \; x \in R^d \;, \tag{6.22}$$

where we also write $A_\alpha(f)(x)$ as the indicator function of set $A_\alpha(f)$.

Thus, the pointwise estimation of $f(x)$ can be achieve by estimating each $\alpha-level$ set of f, say by some random set $A_{\alpha,n}$ (based upon an i.i.d. sample $X_1, X_2, ..., X_n$), and using a plug-in estimator of the form

$$f_n(x, X_1, X_2, ..., X_n) = \int_0^\infty A_{\alpha,n}(x)d\alpha \,. \tag{6.23}$$

Of course, it remains to examine whether such an estimator possesses desirable properties.

The method of “excess mass” provides a rationale for proposing random set estimators (see e.g. Hartigan [33], Polonik [70]).

Specifically, let F denote the unknown distribution function of X. We write $F(dx)$ for associated Stieltjes probability measure, i.e. the probability laws of X on $\mathcal{B}(\mathcal{R}^d)$. Let $L(dx)$ denote the Lebesgue measure on $\mathcal{R}^d$. Then for each $A \in \mathcal{B}(\mathcal{R}^d)$, and $\alpha \in R^+$, $\varepsilon_\alpha(A) = (F - \alpha L)(A)$ is the excess mass of the set A at level α. By writing $A = (A \cap A_\alpha) \cup (A \cap A_\alpha^c)$, we see that

$$\varepsilon_\alpha(A) \leq \varepsilon_\alpha(A_\alpha), \; \forall A \in \mathcal{B}(\mathcal{R}^d),$$

i.e. the level set A_α has the largest excess mass at level α among all *Borel sets*.

This suggest a way to estimate A_α from empirical counter-part of the signed measure $F - \alpha L$, just like the maximum likelihood principle in statistics.

Let $F_n(dx)$ denote the empirical measure associated with the sample $X_1, X_2, ..., X_n$, i.e.

$$F_n(dx) = \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j} \;, \tag{6.24}$$

where δ_X being the Dirac measure as $X \in \mathcal{R}^d$. The empirical excess mass of $A \in \mathcal{B}(\mathcal{R}^d)$ is then

$$\varepsilon_{\alpha,n}(A) = (F_n - \alpha L)(A) . \tag{6.25}$$

Then, formally, an "*excess mass*" estimator for A_α could be taken as $A_{\alpha,n}$ which is the set maximizing $\varepsilon_{\alpha,n}(A)$ over some specified class C of subsets of $\mathcal{B}(\mathcal{R}^d)$.

While the program seems plausible, much more work is needed to justify its use in statistical practice, e.g. *limit distributions* of the random set estimator $A_{\alpha,n}$. This requires, among other things, tractable criteria for convergence in distribution of sequences of random sets. Moreover, the above optimization problem needs special attention: the variable A is neither a vector, nor a function, but a set. How to maximize $\varepsilon_{\alpha,n}(A)$ over $A \in C$?

From a general view point, this boils down to develop a *variational calculus of set functions*.

While the above type of optimization of set functions occurs in many areas of applied mathematics, such as shape optimization, we are not aware of set functions for solving it. We propose below a *variational calculus* (see Nguyen and Kreinovich [59]).

But first, here is some familiar situations in statistics where optimization of set functions arises.

(i) *Neyman-Pearson lemma*

Let X be a random vector on R^d. Consider simple hypotheses

H_0 *: The probability law of X is the probability measure* P_0*,*

H_a *: The probability law of X is* P_a.

Suppose P_0, P_a can be represented in terms of densities (f_0, f_a) with respect to a common measure L on $\mathcal{B}(\mathcal{R}^d)$.

The *type-I* error is

$$\alpha = P(x \in B | H_a) = \int_B f_o(x) dL(x), \tag{6.26}$$

where B is the critical region of a test.

The *type-II error* is

$$\beta = P(x \notin B | H_a) = \int_{B^c} f_a(x) dL(x). \tag{6.27}$$

For fixed α, the problem of finding B to minimize β is this.

Find B_* in $\tau = \{B \in \mathcal{B}(\mathcal{R}^d) : \int_B f_o(x) dL(x) \le \alpha\}$ so that

$$\int_{B^*} f_0(x) dL(x) = \alpha \tag{6.28}$$

and

$$\int_{B_*^c} f_a(x) dL(x) \le \int_{B^c} f_a(x) dL(x) \text{ for all } B \in \tau. \tag{6.29}$$

Neyman and Pearson derived the solution of this optimization problem of set functions using the calculus of variations.

The solution is

$$B_* = \{x \in R^d : f_a > k f_o(x)\} \tag{6.30}$$

where k is such that $\int_{B_*} f_0(x) dL(x) = \alpha$.

(ii) *Bayes tests between capacities* (see Huber and Strassen [37])

Consider the following situation in robust statistics. Let M denote the set of all probability measures on the complete, separable and metrizalbe space Ω. Consider testing thee null hypothesis H_0 against the alternative H_a, where

$$H_0 = \{P \in M : P \le T_o\}$$

$$H_a = \{P \in M : P \le T_a\},$$

where T_0, T_a are two capacity functionals.

Let B be a critical region for the above test, i.e. rejecting H_0 if $x \in B$. is observed. Then the upper probability of falsely rejecting H_0 is $T_0(B)$, and of falsely rejecting H_a is $T_a(B^c)$.

If we assume that H_0 is true with prior probability $\frac{t}{1+t}$, $0 \le t < \infty$, then the (upper) Bayes risk of B is

$$\frac{1}{1+t}T_0(B) + \frac{1}{1+t}T_a(B^c) = \frac{1}{1+t}T_0(B) + \frac{1}{1+t}(1 - V_a(B)). \tag{6.31}$$

Where V_a is the conjugate of T_a, i.e. $V_a = 1 - T_a(B^c)$.

Thus, we are led to minimizing the set function $B \to tT_0(B) - V_a(B)$, for t fixed.

(iii) Dividing a territory

Let U be an open and bounded set of $\mathcal{R}^d$, and $g_i : U \to \mathcal{R}^d$, i=1,2, be continuous functions. Find a closed set $A \subseteq U$ such that

$$(\int_A g_1(x)dx)(\int_{U \setminus A} g_2(x)dx)$$

is maximum over the class $\mathcal{F}(U)$ of closed subsets of U.

As far as we know, there seems to be no appropriate available variational calculus for the above type of problems. As such, below is our attempt to provide a tool for solving optimization problems of set-functionals. We are going to define a concept of derivative DF for a set-function F, defined on some class of subsets of a set U so that when DF(A) = 0, $F(A)$ will be a stationary value for f. Although this can be formulated in a fairly general context, we restrict ourselves to the concrete case of $U = \mathcal{R}^d$, and the domain of F is $B(\mathcal{R}^d)$, i.e. $F : B \to \mathcal{R}$.

In order to extend the classical notion of derivatives of functions to set-functions, we choose to use the symmetric form of derivatives (also called Schwartz derivatives).

For $F:\mathcal{R} \to \mathcal{R}$, defined in some neighborhood of x, and provided that the limit exists, the symmetric (Schwartz) derivative of f at x is by definition

$$f'(x) = \lim_{h \to 0} . \frac{f(x+h) - f(x-h)}{2h} \tag{6.32}$$

Remark 6.1 This type of symmetric derivatives is useful in numerical estimation since its truncation error is of order of $O(h^2)$. Recall that if exists $f'(x) = \lim_{h \to 0} . \frac{f(x+h) - f(x-h)}{2h}$, then $f'(x)$ leads to the same value. However, the converse is not true**:** it is possible that f does not have a derivative in the usual sense, but yet it has a derivative $f'(x)$ above.

Our second idea about extending $f'(x)$ to the setting of set-functions is based upon an analogy with partial derivative of a function $f(x_1, \dots x_n)$ of several variables, we need to specify two things**:** the point $(x_1, \dots x_n)$ at which the derivative is computed, and which variable x_i (among $x_1, \dots, x_n$) over which we differentiate.

Thus, for $F:\mathcal{B} \to \mathcal{R}$, the *analogy* is this. We need to evaluate $1_A(t)$ for all $t \in \mathcal{R}^d$. Viewing the $1_A(t)$'s as "*variables*", we need also to specify a particular variable $1_A(t)$, or equivalently a $t \in \mathrm{R}^d$. Therefore we are led to consider the notion of derivative of F at (A,t) for $A \in \mathcal{B}$, $t \in \mathcal{R}^d$. Sepcifically,

Definition 6.1 *Let $F:\mathcal{B}(\mathcal{R}^d) \to \mathcal{R}$, and τ be the Lebesgue measure on $\mathcal{B}(\mathcal{R}^d)$. The derivative of F at (A,t), for $A \in \mathcal{B}$, $t \in \mathcal{R}^d$ is defined to be*

$$DF(A,t) = \lim_{H \to \{t\}} \frac{F(A \cup H) - F(A \cap H')}{\tau(H)} \tag{6.33}$$

provided the limit exists, where the limit is taken over H such that $\tau(H) \neq 0$, and $H \to \{t\}$ in the Hausdorff metric sense.

Remark 6.2 If $\|\ \|$ denotes the usual norm on $\mathcal{R}^d$, then the (extended, pseudo) Hausdorff metric on non-empty subsets of $\mathcal{R}^d$ is

$$d(A,B) = \max\{\sup_{a\in A}\inf_{b\in B} \| a-b \|, \sup_{b\in B}\inf_{a\in A} \| a-b \|\}.$$

Clearly, the definition of $DF(A,t)$ is inspired from $f'(x)$ in which its numerator $f(x+h)-f(x)$ is replaced by $F(A\bigcup H)-F(A\setminus H)$ whereas its denominator $2h$ (the length of $[-h,h]$) is replaced by $\tau(H)$ with $\tau(H)\neq 0$. ($[-h,h]\to\{0\}$, a singleton, corresponds to $H\to\{t\}$).

Note also that if F is *additive*, then $F(A\bigcup H)-F(A\bigcap H')=F(H)$, and hence the definition of $DF(A,t)$ turns into the so-called "*general derivative*"

$$DF(t) = \lim_{H\to\{t\},\tau(H)\neq 0} \frac{F(H)}{\tau(H)}. \tag{6.34}$$

If F is a measure which is absolutely continuous with respect to τ, then its Radon-Nikodym derivative $\frac{dF}{dt}(t)$ is equal almost everywhere to $DF(t)$.

Definition 6.2 *The set-function $F:B\to\mathcal{R}$ is said to be differentiable if it has derivative at every point $(A,t)\in\mathcal{B}\times\mathcal{R}^d$. F is said to be continuously differentiable if it is differentiable and the map $t\to DF(A,t)$ is continuous for every $A\in B(\mathcal{R}^d)$*

Example 6.1.

(i) In the territorial division problem, the set-functions F_1, F_2, defined on subsets $A\subseteq U$, are closely continuously differentiable, where

$$F_1(A)=\int_A \upsilon_1(x)d\tau(x),\ F_2(A)=\int_U \upsilon_2(x)d\tau(x)-\int_A \upsilon_2(x)d\tau(x)$$

with

$$DF_1(A,t) = \upsilon_1(t)\ ,\ DF_2(A,t) = -\upsilon_2(t)\ .$$

And hence $F(A) = F_1(A)F_2(A)$ is also continuously differentiable with

$$DF(A,t) = \upsilon_1(t)F_2(A) - \upsilon_2(t)F_1(A) \tag{6.35}$$

(ii) The excess mass set-function

$$\varepsilon_\alpha(A) = (F - \alpha\tau)(A)$$

is continuously differentiable if the density $f(x) = \dfrac{dF}{dt}(x)$ is continuous.

As such,

$$D\varepsilon_\alpha(A,t) = f(t) - \alpha\ . \tag{6.36}$$

In the following, the interior, closure and boundary of a set A is denoted as A°, $\overline{A}$, $\delta(A)$, respectively. Of course, we assume $A^\circ \neq \phi$.

Observe that if F attains its maximum at A, then $DF(A,t) \geq 0$ or ≤ 0 for all $t \in A^\circ$ or all $t \in (A')^\circ$ respectively. Indeed assume that A is a set at which F is maximum. For $t \in A^\circ$, the open ball $B(t,r)$ = $\{x \in R^d : \| x - t \| < r\}$ $\subseteq A$ for r sufficiently small. Thus, $F(A \cup B(t,r))$ $= F(A)$ and hence $DF(A,t) \geq 0$ since $F(A)$ is maximum value. Similarly, for $t \in (A')^\circ$, $B(t,r) \subseteq A$ for small γ and hence $A \setminus B(t,r) = A$, implying that $DF(A,t) \leq 0$. The situation for minimum is dual**:** $DF(A,t) \geq 0$ or ≤ 0 according to $t \in A^\circ$ or $t \in (A')^\circ$.

The following result is useful for optimization of set-functions.

Theorem 6.1 *Let F be continuously differentiable. If F attains maximum (or minimum) at some set A (with* A° *and* $(A')^\circ$ *non-empty), then necessarily* $DF(A,t) = 0$ *for all* $t \in \delta(A^\circ) \bigcap \delta((A')^\circ)$.

Proof: Suppose F attains its maximum at A. For $t \in \delta(A^\circ)$, t is limit of a sequence $t_n \in A^\circ$ with $DF(A,t_n) \geq 0$, by continuity. Similarly, for $t \in \delta((A')^\circ)$, t is limit of a sequence $s_n \in (A')^\circ$ with $DF(A,s_n) \leq 0$ for each n, and hence $DF(A,t) \leq 0$.

6.4 Choquet Weak Convergnece

The material in this section (drawn from Feng and Nguyen [53] and from Ding Feng's 2004 Ph.D. Thesis) is about convergence of capacity functionals of random sets. The motivation stems from asymptotic aspects in inference and decision-making with coarse data in Biostatistics, set-valued observations, as well as connections between random sets with several emerging uncertainty calculi in intelligent systems such as fuzziness, belief functions and possibility theory. Specifically, we study the counter-part of Billingsley's Portmanteau Theorem for weak convergence of Choquet integrals.

Recall that a random closed set is a random element in the space of all closed sets of the basic setting space. In the classical theory of random sets, the setting space is assumed to be a *locally compact, second-countable Hausdorff* (LCSCH) space, see e.g. Matheron [ ]. Throughout this section, U is always considered as a LCSCH space. Especially, the Euclidean space $\mathcal{R}^d$ is a nice example of LCSCH spaces. Let $\mathcal{F}$ = {all closed sets in U}, $\mathcal{K}$ = {all compact sets in U}, $\mathcal{G}$ = {all open sets in U}, $\mathcal{K}_0 = \mathcal{K}$- $\{\phi\}$, and $\mathcal{B}(U)$= {all Borel sets in U}.

Let $\mathcal{B}(\mathcal{F})$ be the σ-algebra generated by classes $\mathcal{F}^{\mathcal{K}}$, $K \in \mathcal{K}$ and F_G, $G \in g$, where

$$F^k = \{F \in F : F \bigcap K \neq \phi\}, \ F_G = \{F \in F : F \bigcap G \neq \phi\} \tag{6.37}$$

Define a random closed set as a measurable mapping $S : (\Omega, \Sigma, P) \to$ ($\mathcal{F}$, $\mathcal{B}(\mathcal{F})$) from a probability space into a measurable space. Let Q be the probability measure induced on $\mathcal{B}(\mathcal{F})$, that is, for each $B \in \mathcal{B}(\mathcal{F})$, $Q(B) = P(S^{-1}(B))$.

Note that the distribution of random set S is uniquely determined by its capacity functional T_S on $\mathcal{K}$ defined as:

$$T_S(K) = P(\{\omega : S(\omega) \bigcap K \neq \phi\}), \quad K \in \mathcal{K}$$

It is easy to prove that T_S defined above satisfies the following properties:

(T1) T_S is upper semi-continuous (u.s.c) on $\mathcal{K}$, i.e.,

$$K_n \downarrow K \text{ in } \mathcal{K} \rightarrow T_S(K_n) \downarrow T_S(K);$$

(T2) $T_S(\phi) = 0$ and $0 \leq T_S \leq 1$;

(T3) T_S is monotone increasing on $\mathcal{K}$ and for

$$K_1, K_2, \ldots, K_n \in \mathcal{K}, \ n \geq 2,$$

$$T_S(\bigcap_{i=1}^{n} K_i) \leq \sum_{\phi \neq I \subset (1,2,\ldots,n)} (-1)^{|I|+1} T_S(\bigcup_{i \in I} K_i).$$

Let Γ denote the class of all the functionals T on $\mathcal{K}$ satisfying (T1)-(T3) above. Such fuctionals are called alternating Choquet capacities of infinite order or, for brevity, simply *capacity functionals.*

In the LCSCH space U, the capacity functional T can be extended to Borel sets by

$$T(B) = \sup\{T(K) : K \in \mathcal{K}, K \subset B\} \text{ for } B \in \mathcal{B}(U)$$

Note that

(T1') Upper semi-continuity of T on g means that

$$G_n \uparrow G \text{ in } \mathcal{G} \Rightarrow T_S(G_n) \uparrow T_S(G);$$

We will use the following lemma in our main results.

Lemma 6.1 (*Choquet's Theorem*). *Every probability measure Q on* ($\mathcal{F}$, $\mathcal{B}(\mathcal{F})$ *determines a Choquet capacity functional T on K through the correspondences*

(C1). $T(K) = Q(\mathcal{F}_K)$, $\forall K \in \mathcal{K}$, *and*

(C2). $T(G) = Q(F_G)$, $\forall G \in \mathcal{G}$.

Conversely, ever Choquet capacity functional T on κ *determines a unique probability measure Q on* ($\mathcal{F}$, $\mathcal{B}(\mathcal{F})$ *that satisfies the above conditions (C1) and (C2).*

A compact set $K \in \mathcal{K}$ is called a continuity set for T if and only if $T(K) = T(K^o))$ where K^o denotes the interior of K. Thus a compact set K is T-continuous if and only if $\mathcal{F}_K$ is Q-continuous. Let

$$S_T = \{K \in \mathcal{K}: T(K) = T(K^o)\}.$$

Results on the weak convergence of random closed sets in U have been discussed in many papers, see, e.g., Salinetti and Wets [76], Norberg [65], and Molchanov [53]. The problem we are going to address is this. Since the analysis of random sets is much simpler in terms of their capacity functionals T_n, T, it is desirable to establish the convergence in distribution in term of the T_n's and T, rather than the Q_n's and Q. The following lemma is given by Molchanov [54].

Lemma 6.2 *Let U be a LCSCH space. Then the following statements are equivalent:*

(a). $Q_n \xrightarrow{W} Q$;

(b). $\lim_n T_n(A) = T(A), \forall A \in S_\gamma$.

In order to study the counter-part of Billingsley's Portmanteau Theorem, see Billingsley [31]. We need first an appropriate definition of weak convergence for capacity functionals (which are generalizations of probability measures on $(U, \mathcal{B}(U)$. Let $C_b(U)$ be the space of all bounded real-valued continuous functions on U. For $f \in C_b(U)$, the *Choquet integral* of f with respect to T is

$$\int f dT := \int_0^{+\infty} T(\{f \geq t\})dt + \int_{-\infty}^{0} [T(\{f \geq t\}) - T(U)]dt . \tag{6.38}$$

We say that the sequence of capacity functionals T_n converges, *in the Choquet weak sense,* to the capacity T in U, if

$$\int f dT_n \to \int f dT,\ \forall f \in C_b(U),\ \text{in symbol, } T_n \xrightarrow{C-W} T \text{, as } n \to \infty .$$

Note that the capacity functional T on the setting space U has the following properties (see, Matheron (1975)):

(a) $T(G) = \sup\{T(K) : K \in \mathcal{K}, K \subset G\}$, $G \in \mathcal{G}$;

(b) $T(K) = \inf\{T(G) : G \in \mathcal{G},\ G \supset K\}$, $K \in \mathcal{K}$.

Let τ_f denote the *hit-or-miss topology* on $\mathcal{F}$ generated by two classes, $\mathcal{F}^K, K \in \mathcal{K}$, and $\mathcal{F}_G$, $G \in \mathcal{G}$. Now observe that $\mathcal{F}$, with the hit-or-miss topology, is *compact and metrizable*. Thus, the convergence in distribution of a sequence of random closed sets S_n to S can be studied in the weak topology on this metric space.

A subset $B \subset U$ is called *T-continuous functional closed set* if $B = \{x : f(x) \geq a\}$ for some $a \in \mathcal{R}$, $f \in C_b(U)$ such that

$$T(\{x : f(x) \geq a\}) = T(\{x : f(x) > a\}) . \tag{6.39}$$

For a given capacity T, a family of capacity functionals $A \subset \Gamma$ is said to be *T-tight* if for any *T-continuous functional closed set* B and for each $\varepsilon > 0$, there exists a compact set $K_B \subset B$ such that

$$\sup_{R\in A}[R(B)-R(B\bigcap K_B)]<\varepsilon\,. \tag{6.40}$$

Especially, since U is *T-continuous functional closed set* for any give *T*, there exists a compact set $K_U \subset U$ such that

$$\sup_{R\in A}[R(U)-R(K_U)]<\varepsilon\,. \tag{6.41}$$

Lemma 6.3 *Let* $T\in\Gamma$ *and* $f\in C_b(U)$. *Then*

$$\int_a^b T(\{f\geq t\})dt=\int_a^b T(\{f>t\})dt, \forall a,b\in\Re\,.$$

Moreover , we have

$$\int f dT=\int_{-\|f\|}^{\|f\|}T(\{f\geq t\})dt-T(U)\|f\|$$

$$=\int_{-\|f\|}^{\|f\|}T(\{f>t\})dt-T(U)\|f\|\,.$$

Proof. To see these, simply note that

$$\int_a^b T(\{f\geq t\})dt=\lim_{n\to\infty}\int_{a+\frac{1}{n}}^{b+\frac{1}{n}}T(\{f\geq t\})dt$$

$$=\lim_{n\to\infty}\int_a^b T(\{f\geq t+\tfrac{1}{n}\})dt\leq\int_a^b T(\{f>t\})dt$$

$$\leq\int_a^b T(\{f\geq t\})dt\,. \tag{6.42}$$

Thus $\int_a^b T(\{f\geq t\})dt=\int_a^b T(\{f>t\})dt$. (6.43)

From this lemma, we know that $\{f\geq t\}$ is *T-continuous functional closed set* almost everywhere on $t\in(-\|f\|-1,\|f\|+1)$.

Put $W=\{B \in \mathcal{B}(U) : \overline{B} \in \mathcal{K}\}$. Say that a class $L \subset W$ is *separating* if for any $K \in \mathcal{K}$ and $G \in \mathcal{G}$ with $K \subset G$, there exists some $A \in \mathcal{L}$ such that $K \subset A \subset G$. The following proposition is a conclusion of Theorem 2,1 in Norberg (65).

Proposition 6.1 *Suppose U is a LCSCH space. Let S_1 S_2 ,….be random closed sets in U, T_1, T_2,…their corresponding capacity functional, and Q_1 Q_2,….their corresponding probability distributions. If there exists a separating class $L \subset W$ and a capacity functional R such that*

$$R(A^o) \leq \liminf_{n\to\infty} T_n(A) \leq \limsup_{n\to\infty} T_n(A) \leq R(\overline{A}) \text{ , } \forall A \in L \text{ .}$$

then there exists a random set S in U with the probability distributions Q and hitting functional T satisfying $Q_n \xrightarrow{W} Q$ and $T = R$ on $\mathcal{K}$.

The main result presented here is that, for the case where U is LCSCH, $T_n \xrightarrow{C-W} T$ is equivalent to the convergence in distribution of corresponding random sets with T-tightness of the sequence $\{T_n\}_{n=1}^{\infty}$. Let Q_n, Q denote probability measures on $\mathcal{B}(\mathcal{F})$, and T_n, T denote corresponding capacity functional.

Example 6.2

Let $U = \mathcal{R}$ Take F_n:$\mathcal{R}$ →[0,1] defined by

$$F_n(x) = \begin{cases} x^n \\ 1 + x \end{cases}$$

Then $F_n(x)$ is the distribution of a random varible X_n. Since $F_n(x) \to 0$ for each x $\mathcal{R}$, X_n is not weakly convergent. Let T_n(K) be the capacity functional of random closed set S_n={X_n}. It is easy to see $T_n(K) \to 0$, $\forall K \in \mathcal{K}$ by

$$0 \leq T_n(K) \leq F_n(\max(K)) \to 0$$

Let Q_n denote probability measures on $\mathcal{B}(\mathcal{F})$ determined by T_n and Q be probability measure on $\mathcal{B}(\mathcal{F})$ determined by the capacity 0. Since $T_n(K) \to 0$, $\forall K \in \mathcal{K}$, we get $Q_n \xrightarrow{W} Q$ by the above Lemma B but $T_n \xrightarrow{C-W} T$ does not hold.

Indeed, let $f(x)$=1. Then for 0<t<1, we have

$$T_n(\{f \geq t\}) = 1$$

So we obtain

$$\int f dT = \int_0^1 T_n(\{f \geq t\})dt = 1$$

and

$$\int f d0 = \int_0^1 0(\{f \geq t\})dt = 0$$

Thus $T_n \xrightarrow{C-W} 0$ does not hold.

Lemma 6.4

If $T_n \xrightarrow{C-W} T$, then $\limsup_{n\to\infty} T_n(B) \leq T(B)$, $\forall B \in \mathcal{F}$

Proof: Let $\phi \neq B \in F$. By the definition of the capacity T (e.g., see page 30 in Matheron []) there exists a decreasing sequence $\{G_m\} \subseteq \mathcal{G}$ such that $G_m \searrow B$ and $T(G_m) \searrow T(B)$. Hence B and $U \searrow G_m$ are disjoint closed sets in U. By Urysohn's lemma, there exists a continuous function: $U \to [0,1]$ such that $f(x) = 0$ for $x \in U \backslash G_m$, and $f(x)$=1 for $x \in \mathcal{B}$. Note that

$$\{f \ > t\} = \phi \text{ for } t \geq 1, \ \ \{f \ > t\} = U \text{ for } t < 0$$

Moreover for $0 \leq t < 1$, it implied

$$B \subset \{f_m > t\} \subset \{f_m > 0\} \subset G_m . \tag{3}$$

By Lemma 6.3 we know

$$\int f_m dT_n = \int_0^1 T_n(\{f_m > t\})dt \text{ , } n = 1,2,\cdots .$$

And

$$\int f_m dT = \int_0^1 T(\{f_m > t\})dt$$

Since $T_n \xrightarrow{C-W} T_\infty$, we obtain

$$\limsup_{n\to\infty} T_n(B) = \limsup_{n\to\infty} \int_0^1 T_n(B)dt$$

$$\leq \limsup_{n\to\infty} \int_0^1 T_n(f_m > t)dt$$

$$= \limsup_{n\to\infty} \int f_m dT_n = \int f_m dT$$

$$= \int_0^1 T(\{f_m > t\})dt \tag{6.44}$$

By using (3) we get

$$\limsup_{n\to\infty} T_n(B) \leq \liminf_{n\to\infty} \int_0^1 T(\{f_m > t\})dt$$

$$\leq \lim_{n\to\infty} \int_0^1 T(G_m)dt = T(B) \tag{6.45}$$

Lemma 6.5

If $T_n \xrightarrow{C-W} T$ *, then* $\liminf_{n\to\infty} T_n(G) \geq T(G)$ *,* $\forall G \in \mathcal{G}$*.*

Proof: Let $\phi \neq G \in g$. By the property of the capacity functional T (e.g., see Matheron [51]), there exists a increasing sequence $\{K_m\} \subset \mathcal{K}$ such that $K_m \uparrow G$ and $T(K_m) \uparrow T(G)$. Hence K_m and $U \downarrow G$ are disjoint closed sets in U. By Urysohn's lemma, there exists a continuous function

$g_m : U \to [0,1]$ such that $g_m(x) = 0$ for ** $U \downarrow G$, and $g_m(x) = 1$ for $x \in K_m$. Simply note that $\{g_m \geq t\} = \phi$ for t >1, $\{g_m \geq t\} = U$ for $t \leq 0$. Moreover , for $0 < t \leq 1$, it implies

$$G \supset \{g_m > 0\} \supseteq \{g_m \geq t\} \supset K_m \tag{4}$$

All these mean that

$$\int g_m dT_n = \int_0^1 T_n(\{g_m \succ t\})dt \text{ , } n = 1,2,\cdots .$$

And

$$\int g_m dT = \int_0^1 T(\{g_m \geq t\})dt$$

By $T_n \xrightarrow{C-W} T$ and the monotone property of capacity functionals, we obtain

$$\liminf_{n\to\infty} T_n(G) = \liminf_{n\to\infty} \int_0^1 T_n(G)dt \tag{6.46}$$

$$\geq \liminf_{n\to\infty} \int_0^1 T_n(\{g_m \geq t\})dt$$

$$= \liminf_{n\to\aleph} \int g_m dT_n = \int g_m dT = \int_0^1 T(\{g_m \geq t\})dt \tag{6.47}$$

By (4), we imply

$$\liminf_{n\to\infty} T_n(G) \geq \limsup_{n\to\infty} \int_0^1 T_n(\{g_m \geq t\})dt$$

$$\geq \limsup_{m\to\infty} \int_0^1 T(K_m)dt = T(G) \tag{6.48}$$

It is easy to show that any family of finitely many of capacity functionals is T-tight for any given T. By the Lemma C and Lemma D, however, we conclude the following result:

Theorem 6.2 *If* $T_n \xrightarrow{C-W} T$ *, then the sequence* $\{T_n\}$ *is T-tight.*

Proof: Since U is LCSCH, there exists a sequence of open set $\{C_p\}_{q=1}^{\infty}$ such that $\overline{C_p} \in k$, $\forall q \geq 1$ and $C_q \uparrow U$. Let $B = \{x : f(x) \geq a\}$ be a *T-continuous functional closed set* with

$$T(\{x : f(x) \geq a\}) = T(\{x : f(x) > a\}) \tag{6.49}$$

By the upper semi-continuity of *T*, $[C_q \cap \{x : f(x) > a\}] \uparrow \{x : f(x) > a\}$ implies

$$T[C_q \cap \{x : f(x) > a\}] \uparrow T\{x : f(x) > a\}$$

For each ε> 0, there exists a q_0 such that

$$0 \leq T(\{x : f(x) > a\}) - T(C_q \cap \{x : f(x) > a\}) < \tfrac{\varepsilon}{2}, \forall q \geq q_0$$

By Lemma 6.3 and Lemma 6.4, we have

$$\limsup_{n\to\infty}[T_n(B) - T_n(\overline{C_q} \cap B)]$$

$$\leq \limsup_{n\to\infty}[T_n(B) - T_n(\overline{C_q} \cap \{x : f(x) > a\})]$$

$$\leq \limsup_{n\to\infty}[T_n(B) - \liminf_{n\to\infty} T_n(C_q \cap \{x : f(x) > a\})]$$

$$\leq T(B) - T(C_q \cap \{x : f(x) > a\})$$

$$= T(\{x : f(x) > a\}) - T(C_q \cap \{x : f(x) > a\}) < \tfrac{\varepsilon}{2} \tag{6.50}$$

Since any family of finitely many of capacity functionals is T-tight, we obtain the sequence $\{T_n\}$ is T-tight.

Lemma 6.5 *If* $Q_n \xrightarrow{C-W} Q$ *and the sequence* $\{T_n\}_{n=1}^{\infty}$ *is T-tight, then*

$$\liminf_{n\to\infty} \int f dT_n \geq \int f dT \text{ , } \forall f \in C_b(U) \text{ .}$$

Proof: For any given $f \in C_b(U)$, Lemma 6.4 implies

$$\int f dT_n = \int_{-\|f\|}^{\|f\|} T_n(\{f > t\})dt - T_n(U) \| f \| \text{ , } \forall n \geq 1$$

$$\int f dT = \int_{-\|f\|}^{\|f\|} T_n(\{f > t\})dt - T_n(U) \| f \| \tag{6.51}$$

Note that $F_{\{f \geq t\}}$ is open in the $\mathcal{F}$, with the hit-or-miss topology. Since $Q_n \xrightarrow{W} Q$, Billingsley's Portmanteau Theorem (e.g. see Billingsley [5] implies

$$\liminf_{n\to\infty} Q_n(F_{\{f \geq t\}}) \geq Q(F_{\{f > t\}}) \text{ , } \forall t \in (-\| f \|, \| f \|) \tag{6.52}$$

On the other hand, for each ε > 0, by the T-tightness of the sequence $\{T_n\}_{n=1}^{\infty}$, there exists a compact $K \subset U$ such that

$$\sup_n [T_n - T_n(K)] < \varepsilon$$

Simply note that $\mathcal{F}_K$ is closed in the $\mathcal{F}$, with hit-or-miss topology. For $Q_n \xrightarrow{W} Q$, by Billingsley's Portmanteau Theorem, we obtain $\sup_{n\to\infty} Q_n(F_k) \leq Q(F_k)$. By Choquet's Theorem and Fatou's Lemma, we obtain

$$\liminf_{n\to\infty} \int f dT_n$$

$$\geq \liminf_{n\to\infty} \int_{-\|f\|}^{\|f\|} T_n(\{f > t\})dt - \limsup_{n\to\infty} T_n(U) \| f \|$$

$$\geq \liminf_{n\to\infty} \int_{-\|f\|}^{\|f\|} Q_n(F_{\{f>t\}})dt - \limsup_{n\to\infty} [T_n(K+\varepsilon)] \| f \|$$

$$\geq \int_{-\|f\|}^{\|f\|} \liminf_{n\to\infty} Q_n(F_{\{f>t\}})dt - \limsup_{n\to\infty} Q_n(F_K) \| f \| - \varepsilon \| f \|$$

$$\geq \int_{-\|f\|}^{\|f\|} Q(F_{\{f>t\}})dt - Q(F_K) \| f \| - \varepsilon \| f \| \tag{6.53}$$

i.e.

$$\liminf_{n\to\infty} \int f dT_n$$

$$\geq \int_{-\|f\|}^{\|f\|} Q(F_{\{f>t\}})dt - Q(F_U) \| f \| - \varepsilon \| f \|$$

$$= \int_{-\|f\|}^{\|f\|} T(f > t)dt - T(U) \| f \| - \varepsilon \| f \|$$

$$= \int f dt - \varepsilon \| f \|$$

Since ε is arbitrary, it implies

$$\liminf_{n\to\infty} \int f dT_n \geq \int f dT \quad . \tag{6.54}$$

Lemma 6.6 *If $Q_n \xrightarrow{W} Q$ and the sequence $\{T_n\}_{n=1}^{\infty}$ is T-tight, then*

$$\limsup_{n\to\infty} \int f dT_n \leq \int f dT \text{ , } \forall f \in C_b(U) .$$

Proof: For any fixed $f \in C_b(U)$, the decreasing function $T(\{f > t\})$ is Riemann integrable on $[-\| f \|, \| f \|]$. By Lemma B, $(\{f \geq t\})$ is T-

continuous functional closed set almost everywhere on $t \in (-\| f \| -1, \| f \| +1)$. For each ε> 0, there exists a subdivision of

$$-\| f \| = t_0 < t_1 < t_2 < \cdots < t_j < t_{j+1} < \cdots < t_m = \| f \|$$

such that

(a) $\{f \geq t_j\}$ is T-continuous functional closed set for j=0,1,2…m-1;

(b) the following inequality holds

$$\sum_{j=1}^{m-1} T\{f \geq t_j\}\{t_{j+1} - t_j\} - T(U) \| f \| < \int f dT + \varepsilon \tag{6.55}$$

Since the sequence $\{T_n\}_{n=1}^{\infty}$ is T-tight, there exists a compact set $K \subset U$ such that

$$\sup_n T_n(\{f \geq t_j\}) - T_n(\{f \geq t_j\} \cap K) < \varepsilon \, , \, j = 0,1,2,\ldots,m-1$$

Note that $\mathcal{F}_u$ is open and $F_{(\{f \geq t_j\} \cap K)}$ is closed in $\mathcal{F}$, with the hit-or-miss topology. For $Q_n \xrightarrow{W} Q$, by Bilingsley's Portmanteau Theorem, we obtain

$$\liminf_{n\to\infty} Q_n(F_U) \geq Q(F_U) \, , \, \liminf_{n\to\infty} Q_n(F_{(\{f \geq t_j\} \cap K)}) \leq Q(F_{(\{f \geq t_j\} \cap K)})$$

Hence, monotone increasing of each T_n and Choquet's theorem imply

$$\limsup_{n\to\infty} \int f dT_n$$

$$= \limsup_{n\to\infty} [\int_{-\|f\|}^{\|f\|} T_n(\{f > t\}) dt - T_n(U) \| f \|]$$

$$\leq \limsup_{n\to\infty} [\sum_{j=0}^{m-1} \int_j^{j+1} T_n(\{f \geq t\}) dt] - \liminf_{n\to\infty} Q_n(F_U) \| f \|$$

$$\leq \limsup_{n\to\infty} \sum_{j=0}^{m-1} T_n(\{f \geq t_j\})(t_{j+1} - t_j) - Q(F_U) \| f \|$$

$$\leq \limsup_{n\to\infty} [\sum_{j=0}^{m-1} T_n(\{f \geq t_j\} \cap K + \varepsilon)(t_{j+1} - t_j) - T(U) \| f \|$$

$$\leq \limsup_{n\to\infty} \sum_{j=0}^{m-1} [Q_n(F_{(\{f\geq t_j\}\cap K)}](t_{j+1} - t_j) + 2\varepsilon \| f \| - T(U) \| f \|$$

$$\leq \sum_{j=0}^{m-1} Q(F_{(\{f\geq t_j\}\cap K)})(t_{j+1} - t_j) + 2\varepsilon \| f \| - T(U) \| f \|$$

$$= \sum_{j=0}^{m-1} T(\{f \geq t_j\} \cap K)(t_{j+1} - t_j) + 2\varepsilon \| f \| - T(U) \| f \| \qquad (6.56)$$

So we obtain

$$\leq \sum_{j=0}^{m-1} T(\{f \geq t_j\} \cap K)(t_{j+1} - t_j) + 2\varepsilon \| f \| - T(U) \| f \| \qquad (6.57)$$

$$\leq \sum_{j=0}^{m-1} T(\{f \geq t_j\})(t_{j+1} - t_j) - T(U) \| f \| + 2\varepsilon \| f \|$$

$$< \int f dT + \varepsilon + 2\varepsilon \| f \|.$$

Since ε is arbitrary, we obtain $\limsup_{n\to\infty} \int f dT_n \leq \int f dT$.

Using Lemma E and Lemma F, we obtain

Theorem 6.3 *If* $Q_n \xrightarrow{W} Q$ *and the sequence* $\{T_n\}_{n=1}^{\infty}$ *is T-tight, then* $T_n \xrightarrow{C-W} T$.

It turns out that the converse of Theorem 6.3 holds, since

Lemma 6.7 *If* $T_n \xrightarrow{C-W} T$ *, then* $Q_n \xrightarrow{W} Q$.

Proof: We will, at first , prove that the class $S_r \subset W$ is separating. In fact, let $K \in \mathcal{K}$ and $G \in \mathcal{G}$ with $K \subset G$. Since U is LCSCH space, by the compactness of $\mathcal{K}$, there exists an open set $O \subset U$ such that $\overline{O} \in \mathcal{K}$ and $K \subset O \subset G$. Note that K and $U \backslash O$ are disjoint closed sets in U .By Urysohn's lemma, there exists a continuous function $f : U \to [0,1]$ such that $f(x) = 0$ for $x \in U \setminus O$ and $f(x)=1$ for $x \in \mathcal{K}$. Simply note that

$$K \subset \{f \geq t\} \subset O \subset G \, , \forall t \in (0,1) \tag{*}$$

By lemma B, $T\{f \geq t\} = T\{f > t\}$ is almost everywhere on $t \in (0,1)$

Choose $B = \{f \geq t_0\}$ for some $t_0 \in (0,1)$ with

$$T\{f \geq t_0\} = T\{f > t_0\}$$

Keeping in mind that $\{f > t_0\} \subset B^o \subset B$. We get $T(B) = T(Int(B))$ $= T(\{f > t_0\}$. Moreover, $B = \overline{B} \subset \overline{O} \in \mathcal{K}$ implies $B \in \mathcal{K}$ and therefore $B \in S_T$. Thus the class $S_T \subset W$ is separating by (*).

On the other hand, by Lemma C and Lemma D, $T_n \xrightarrow{C-W} T$ implies, for any

$$\limsup_{n\to\infty} T_n(K) \leq T(K)$$

$$\liminf_{n\to\infty} T_n(K^o) \geq T(K^o)$$

Thus, for $K \in S_T$

$$T(K^o) < \liminf_{n\to\infty} T_n(K^o) < \liminf_{n\to\infty} T_n(K) < \limsup_{n\to\infty} T_n(K) < T(K)$$

By the Proposition 6.1, there exists a random set S' with the probability distribution P and the corresponding capacity functional R satisfying $Q_n \xrightarrow{W} P$ and $R = T$ on $\mathcal{K}$. By Choquet's Theorem, we have $Q = P$, i.e. $Q_n \xrightarrow{W} Q$

Theorem 6.4

Let U be a LCSCH space. Then the following statements are equivalent:

(a) $T_n \xrightarrow{C-W} T$ *;*

(b) $Q_n \xrightarrow{W} Q$ *and the sequence* $\{T_n\}_{n=1}^{\infty}$ *is T-tight;*

(c) $\overline{\lim}_n T_n(K) \le T(K)$ *for every compact set* $\mathcal{K}$ *in U,* $\underline{\lim}_n T_n(G) \ge T(G)$ *for every open set* $\mathcal{G}$ *in U and the sequence* $\{T_n\}_{n=1}^{\infty}$ *is T-tight;*

(d) $\underline{\lim}_n T_n(A) = T(A)$ *for each* $A \in S_T$ *, and the sequence* $\{T_n\}_{n=1}^{\infty}$ *is T-tight.*

Corollary 6.1 *Let* $\{U, d\}$ *be a compact metric space. Then the following statements are equivalent*

(a) $T_n \xrightarrow{C-W} T$ *;*

(b) $Q_n \longrightarrow Q$ *;*

(c) $\overline{\lim}_n T_n(K) \le T(K)$ *for every compact set* $\mathcal{K}$ *in U and* $\underline{\lim}_n T_n(G) \ge T(G)$ *for every open set* $\mathcal{G}$ *in U;*

(d) $\lim_n T_n(A) = T(A)$ *for each* $A \in S_T$ *.*

Let $C_S^+(U)$ denote the space of all nonnegative real-valued, continuous function on U with compact supports.

Lemma 6.8 *Let U be a LCSCH space. If* $Q_n \xrightarrow{W} Q$ *, then*

$$\lim_{n\to\infty} \int f dT_n = \int f dT \text{ , } \forall f \in C_S^+(U) .$$

Proof: For any given $f \in C_S^+(U)$, by definition, we know that $\{x \in U : f(x) \neq 0\} \subset K$, For some $K \in \mathcal{K}$. Simply note that Lemma B implies

$$\int f dT_n = \int_0^{\|f\|} T_n(\{f \geq t\})dt = \int_0^{\|f\|} T_n(\{f > t\})dt \qquad (6.58)$$

$$\int f dT = \int_0^{\|f\|} T_n(\{f \geq t\})dt = \int_0^{\|f\|} T(\{f > t\})dt \qquad (6.59)$$

For each $t \in (0, \|f\|)$, the closed set $\{f \geq t\} \subset K$ is compact in the LCSCH space U, so the set $F_{\{f \geq t\}} = F - F^{\{f \geq t\}} =$ is closed in the space $\mathcal{F}$ with the hit-or-miss topology. Moreover $F^{\{f \geq t\}}$ if open in the space $\mathcal{F}$. On the other hand, we know

$$\partial(F_{\{f \geq t\}}) = F_{\{f \geq t\}} - \operatorname{int}(F_{\{f \geq t\}}) \subset F_{\{f \geq t\}} - F_{\{f > t\}} \qquad (6.60)$$

It follows from Choquet's Theorem that

$$0 \leq \int_0^{\|f\|} Q[\partial(F_{\{f \geq t\}})]dt \leq \int_0^{\|f\|} Q(F_{\{f \geq t\}} - F_{\{f > t\}})dt$$

$$= \int_0^{\|f\|} \{(f \geq t)\} - T(\{f > t\})dt = 0$$

All these mean that $F_{\{f \geq t\}}$ is a G-continuous set almost everywhere on $t \in (0, \|f\|)$ .Note the facts $|Q_n| \leq 1$, and the given condition $Q_n \xrightarrow{W} Q$. By using Lebesgue dominated convergence theorem and Billingsley's Portmanteau theorem (see, Billingsley [5]), we imply

$$\lim_{n\to\infty}[\int fdT_n - \int fdT] = \lim_{n\to\infty} \int_0^{\|f\|}[T_n(\{f \geq t\}) - T(\{f \geq t\})]dt$$

$$= \lim_{n\to\infty} \int_0^M [Q_n(F_{\{f\geq t\}}) - Q(F_{\{f\geq t\}})]dt$$

$$= \int_0^M \lim_{n\to\infty}[Q_n(F_{\{f\geq t\}}) - Q(F_{\{f\geq t\}})]dt = 0 \tag{6.61}$$

Let $U=(R^m, d)$ be the Euclidean space. For any $K \in k_0$ j=1,2,…,define

$$\Psi_k^j : R^m \to [0,1] \text{ by } \Psi_k^j(x) := [1 - jd(x,k)] \vee 0 .$$

Then $\Psi_k^j \in C_s^+(R^m)$

Theorem 6.5 *Let* $U = (R^m, d)$. *Then the following statements are equivalent:*

(a) $Q_n \xrightarrow{W} Q$ *;*

(b) $\lim_{n\to\infty} \int fdT_n = \int fdT, \forall f \in C_s^+(U)$

(c) $\lim_{n\to\infty} \int \Psi_K^j dT_n = \int \Psi_K^j dT, \forall K \in k_0, j = 1,2,\ldots$

(d) $\lim_{n\to\infty} \int \Psi_K^j dT_n = \int \Psi_K^j dT, \forall K \in S_T \setminus \{\phi\}, j = 1,2,\ldots$

(e) $\overline{\lim}\, T_n(K) \leq T(K)$ *for every compact set K in U and*

$\underline{\lim}\, T_n(G) \geq T(G)$ *for every open set G in U;*

(f) $\lim_n T_n(K) = T(K)$ *for each* $K \in S_T$.

Proof. From Lemma H we obtain (a)$\to$ (b). The proof of (b) $\to$ (c) and (e) $\to$ (f) are trivial. By Billingsley's Portmanteau theorem (e.g. see, Billingsley (1968)), we can directly get (a) $\to$ (e). The proof of (f) $\to$ (a) is by Lemma B. Now we only need to show (d) $\to$ (a).

In fact, suppose that

$$\lim_{n\to\infty} \int \Psi_K^j dT_n = \int \Psi_K^j dT, \forall K \in S_T \setminus \{\phi\}, j = 1,2,\ldots$$

Since the space $\mathcal{F}$, with the hit-or-miss topology, is a compact metric space, the space of all Borel probability measure is weakly compact. Hence for each subsequence $\{Q_{n'}\}_{n'=1}^{\infty}$ of the sequence $\{Q_n\}_{n=1}^{\infty}$,there exists a further subsequence $\{Q_{n'_j}\}_{j=1}^{\infty}$ weakly converging to some Borel probability measure P on $\mathcal{F}$. Let R be the corresponding hitting capacity functional on k determined by P.

By Lemma 6.4.8, we know that $Q_{n'_j} \xrightarrow{W} P$ implies $\lim_{n\to\infty} \int \Psi_K^j dT_n = \int \Psi_K^j dT, \forall K \in S_T \setminus \{\phi\}, j = 1,2,\ldots$ we have the equality $\lim_{n\to\infty} \int \Psi_K^j dT_n = \int \Psi_K^j dT, \forall K \in S_T \cap S_R \setminus \{\phi\}, j = 1,2,\ldots$ By Lemma I below we imply $T = R$ on $\mathcal{K}$. So Choquet's Theorem implies $Q = P$, i.e. $Q_{n'_j} \xrightarrow{W} Q$. Thus we obtain $Q_n \xrightarrow{W} Q$ by Theorem 2.3 in Billingseley [5].

Lemma 6.9 *Let $U = (\mathcal{R}^n, d)$, and T, R be two capacity functionals on k. If*

$\int \Psi_K^j dR = \int \Psi_K^j dT, \forall K \in S_T \cap S_R \setminus \{\phi\}, j = 1,2,\ldots$*, then $T = R$ on $\mathcal{K}$.*

Proof: Firstly, we will prove that $T = R$ on $S_T \cap S_R$. In fact, if $\phi \neq K \in S_T \cap S_R \setminus \{\phi\}$, then

$$K \subset \{x : \Psi_K^j(x) \geq t\} = \{x : \rho(x,K) \leq \frac{1-t}{j}, \forall t \in (0,1)] \tag{6}$$

This means that $\{x : \Psi_K^j(x) \geq t\}$ is compact in U and $\{x : \Psi_K^j(x) \geq t\} \downarrow K$ as $j \to \infty$.

Hence Lebesgue dominated convergence theorem and the upper semi-continuity of R and T, with $\int \Psi_K^j dT = \int \Psi_K^j dR$ imply

$$R(K) = \int_0^1 \lim_{n\to\infty} R(\{\Psi_K^j \geq t\})dt$$

$$= \lim_{j\to\infty} \int \Psi_K^j dR = \lim_{j\to\infty} \int \Psi_K^j dT$$

$$= \int_0^1 \lim_{j\to\infty} T(\{\Psi_K^j \geq t\})dt$$

$$= T(K).$$

Thus $T = R$ on $S_T \cap S_R$. It remains to show that $T = R$ on $\mathcal{K}$.

Indeed, for any $\phi \neq K$, take f: $\mathcal{R}^m \to [0,1]$ defined by $f(x) = [1\text{-}d(x,K)] \vee 0$, $x \in R^m$

Then Lemma 6.4 implies

$$\int_0^1 T(\{x : f(x) \geq t\})dt = \int_0^1 T(\{x : f(x) > t\})$$

This means that $T(\{x : f(x) \geq t\}) = T(\{x : f(x) > t\}$ almost everywhere in $t \in (0,1)$. Thus

$$0 \leq T(\{x : f(x) \geq t\}) - T(Int\{x : f(x) \geq t\})$$

$$\leq T(\{x : f(x) \geq t\}) - T(\{x : f(x) > t\}).$$

Since $\{x : f(x) \geq t\} = \{s : d(x,K) \leq 1 - t\}$ is compact in U, we imply $\{x : f(x) \geq t\} \in S_T$ almost everywhere in $t \in (0,1)$. Similarly, we obtain $\{x : f(x) \geq t\} \in S_R$ almost everywhere in $\{t_i\}_{i=1}^{\infty} \subset (0,1)$ such that $\{x : f(x) \geq t_i\} \in S_T \cap S_R$ almost everywhere in $t \in (0,1)$.

Choose a sequence $t_i \in (0,1)$ such that $\{x: f(x) \geq t_i\} \in S_T \bigcap S_R$ and $t_i \uparrow 1$ as $n \to \infty$, we have $\{x: f(x) \geq t_i\} \downarrow K$ as $i \to \infty$. By previous part of the proof, we obtain that $R\{x: f(x) \geq t_i\} = T\{x: f(x) \geq t_i\}, i \geq 1$. Thus the upper semi-continuity of R and T implies

$$R(K) = \lim_{i \to \infty} R(\{x: f(x) \geq t_i\}) = \lim_{i \to \infty} T(\{x: f(x) \geq t_i\}) = T(K). \tag{6.62}$$

Example 6.3

Let $U = (\mathcal{R}^m, d)$. Suppose $\{S_n\}$is a sequence of i.i.d. random closed sets in U with the same distribution as the random closed set $S: (\Omega, \Sigma, P) \to (\mathcal{F}$, $\mathcal{B}(\mathcal{F}))$. Let Q_s be the probability measure of S and T_S be its associated capacity functional. To estimate T_S, we use the empirical capacity functional $T_n^{\omega}(K)$, defined by

$$T_n^{\omega}(K) = \frac{1}{n}\sum_{i=1}^{n} 1_{K \cap S_i(\omega) \neq \phi}, \ K \in \mathcal{K}, \tag{6.63}$$

where $1_{K \cap S_i(\omega) \neq \phi} = \begin{cases} 1 & K \bigcap S_i(w) \neq \phi \\ 0 & otherwise \end{cases}$.

Let Q_n be the probability measure of S determined by T_n^{ω}. We claim that $Q_n \xrightarrow{W} Q_s$ almost surely (a.s.). To see this, let B be a countable base of U and L denote the class of all finite unions of the family B. Then L is countable. By the strong law of large numbers, $T_n^{\omega}(B) \to T_s(B)$ as $n \to \infty, B \in L$ for each given . Hence the countability of L implies

$$P\{\omega : T_n^{\omega}(B) \to T_s(B), \forall B \in L\} = 1 \tag{6.64}$$

We only need to show the $Q_n \xrightarrow{W} Q_s$ if $T_n(B) \to T_s(B)$.

Indeed, the compactness of U implies that L is a separating class. To complete proof, simply note that $\lim_{n\to\infty} T_n(B) = T(B)$, for $\forall B \in L$.

Hence we obtain, for each $A \in \mathrm{L}$,

$$T_s(A^o) \le T_s(A) = \liminf_n T_n(A) = \lim_n T_n(A)$$

$$= \limsup_n T_n(A) = T_s(A) \le T_s(A^-) \tag{6.65}$$

By the Proposition 6.1, there exists a random set S' with the distrubution P and the corresponding functional T satisfying $Q_n \xrightarrow{W} P$ and T=T_s on $\mathcal{K}$. By Choquet's Theorem, we have $Q_s = P$, i.e. $Q_n \xrightarrow{W} Q_s$.

Chapter 7 Fuzzy Statistical Analysis and Estimation

In social science research, many decisions, evaluations, or purposes of evaluations are done by surveys or questionnaires to seek for people's consensus. The commonly used method is asking people to think in binary logic way from a multiple choice design. However, these processes often ignore the fuzzy, sometimes even ambiguous thinking behavior perceived in human logic and recognition.

In this chapter we will employ fuzzy sets and fuzzy statistical analysis and will formalize the definitions of fuzzy mode, fuzzy median and fuzzy mean as well as investigating of their related properties. We also give some empirical examples to illustrate the techniques and how to analyze fuzzy data. Results show that fuzzy statistics with soft computing are more realistic and reasonable for the social science research. We include also suggested for the further studies.

7.1 The Nature of Fuzzy Samples

Many fundamental statistics, such as mean, median and mode etc, are useful measurements in illustrating some characteristics for the population distribution. These statistical parameters can be quickly computed from a set of data and its basic information has been widely employed in applications. Each statistics has its special application. For example, when we want to investigate people's opinions or consensus on a certain public issues, the using of mode or median will be more proper than that of mean.

However, traditional statistics are reflecting the result from a two-valued logic opinion. To investigate the population, people's opinions or the complexity of a subjective event more accurately, it is suggested that

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 109–128 (2006)
www.springerlink.com

we should use fuzzy logic. Especially, when we want to know the public opinion on the environmental pollution, fuzzy statistics provides a powerful research tool. Moreover, since Zadeh [117] developed fuzzy set theory, its applications are extended to traditional statistical inferences and methods in social sciences, including medical diagnosis or a stock investment system. For example, [4], [8], [22] and [73] demonstrated the approximate reasoning econometric methods one after another.

There are more and more research focusing on the fuzzy statistical analysis and applications in the social science fields, such as [113] proposed fuzzy statistical testing method to discuss the stationarity of Taiwan short-term money demand function; [109] identified the model construction through qualitative simulation; [8], [22], [106], and [111] demonstrated the concepts of fuzzy statistic and applied it to social survey; [112] used fuzzy regression method of coefficient estimation to analyze Taiwan monitoring index of economic. Recently, along with the raise of intelligent knowledge consciousness and soft-computing, many investigators focus on the application of fuzzy set in calculating the human thought or public polls under the uncertain and incomplete condition.

Since mode, mean and median are essential statistics in analyzing the sampling survey. [6] proposed the generalized median by discussion aggregation operators closely related to medians and to propose new types of aggregation operators appropriate, both for the cardinal and ordinal types of information. But they didn't give a realistic example. In this paper we present the definitions of fuzzy mode, fuzzy median and fuzzy mean, related properties, and proofs. Comprehensive examples about EIA with fuzzy mode, fuzzy median and fuzzy mean on human thought analysis are well illustrated. We hope via these new techniques the people's thought can be extracted and known in a more precise understanding. Finally, conclusions as well as suggestions are provided.

Traditional statistics deals with single answer or certain range of the answer through sampling survey, but it has difficulty in reflecting people's incomplete and uncertain thought. On the other hand, these processes often ignore the intriguing and complicated yet sometimes conflicting human logic and feeling. If people can use the membership function to express the degree of their feelings based on their own perception, the result will be closer to their real thought. For instance, when people process a pollution assessment, they classify the distraction into two categories: pollution and non-pollution. This kind of classification is not realistic, since the pollution is a fuzzy concept (degree) and can hardly be justified by the true-false logic. Therefore, to compute the information based on the fuzzy logic should be more reasonable.

The Degree of Human Feeling

In considering the question related with fuzzy property, we consider that the information itself has the uncertainty and fuzzy property. The following notations for fuzzy numbers will be used in the following chapters for simplicity.

Definition 7.1 Fuzzy Numbers

Let U *be an universal set,* $A=\{A_1,A_2,\cdots,A_n\}$ *be the subset of discussion factors in* U *. For any term or statement* X *on* U *, it's membership corresponding to* $\{A_1,A_2,\cdots,A_n\}$ *is* $\{\mu_1(X),\mu_2(X),\cdots,\mu_n(X)\}$ *, here* $u:U\to[0,1]$ *is a real function. Then the fuzzy number of* X *can be written as the following:*

$$\mu_U(X)=\sum_{i=1}^{n}\mu_i(X)I_{A_i}(X) \tag{7.1}$$

where $I_{A_i}(x)=1,$ *if* $x\in A_i; I_{A_i}(x)=0,$ *if* $x\notin A_i$ *..*

If the domain of the universal set is continuous, then the fuzzy number can be written as： $\mu_U(X)=\int_{A_i\subseteq A}\mu_i(X)I_{A_i}(X)$.

Note that, in many writtings, people are used to write a fuzzy number as $\mu_U(X)=\frac{\mu_1(X)}{A_1}+\frac{\mu_2(X)}{A_2}+\cdots+\frac{\mu_n(X)}{A_n}$ (*where* "+" *stands for "or," and "*

$\div$ " *stands for the membership* $\mu_i(X)$ *on* A_i) instead of $\mu_U(X) = \sum_{i=1}^{n} \mu_i(X) I_{A_i}(X)$.

Example 7.1 How many hours do you do exercise for a day?

Let U *be the universe set which can be seen as an integer set,* $A = \{0,1,2,3,4\}$ *be the hours of exercise per day. Then, a fuzzy number of* X = *hours of exercise can be expressed as*

$$\{\mu_0(X) = 0.25, \mu_1(X) = 0.6, \mu_2(X) = 0.1, \mu_3(X) = 0.05, \mu_4(X) = 0\}$$

Then, the fuzzy number for exercise hours per day can be expressed as

$$\mu_A(X) = \frac{0.25}{0} + \frac{0.6}{1} + \frac{0.1}{2} + \frac{0.05}{3} + \frac{0}{4}$$

In the research of social science, the sampling survey is always used to evaluate and understand public opinion on certain issues. The traditional survey forces people to choose one answer from the survey, but it ignores the uncertainty of human thinking. For instance, when people need to choose the answer from the survey which lists five choices including "Very satisfactory," "Satisfactory," "Normal," "Unsatisfactory," "Very unsatisfactory," traditional survey become quite exclusive.

The advantages of evaluation with fuzzy number include: (i) Evaluation process becomes robust and consistent by reducing the degree of subjectivity of the evaluator. (ii) Self-potentiality is highlighted by indicating individual distinctions. (iii) Provide the evaluators with an encouraging, stimulating, self-reliant guide that emphasizes on individual characteristics. While the drawback is that the calculating process will be a little complex than the conventional one.

Example 7.2 *The use of fuzzy numbers in a sampling survey about favorite colors.*

Consider a fuzzy set of liked colors as shown in Table 3.1. Note that in the extreme cases when a degree is given 1 or 0, that is "white" and "black", a standard "yes" and "no" ar e in complement relationship, as in binary logic.

Table 7.2 Favorite colors

Color	A_1	A_2	A_3
X	$\mu_{A_1}(X)$	$\mu_{A_2}(X)$	$\mu_{A_3}(X)$
White	*1*	*0*	*0*
Black	*0*	*0.7*	*0.3*
Red	*0.8*	*0.2*	*0*
Brown	*0*	*1*	*0*
Pink	*0.6*	*0.4*	*0*
Yellow	*0*	*0.3*	*0.7*

Here A_1 represents the semantics of "favorite color", A_2 is "no difference", and A_3 is "dislike colors". Then fuzzy numbers for these three statements can be represented as:

$$\mu_{A_1}(X) = 1I_{white}(X) + 0.3I_{black}(X) + 0.8I_{red}(X) + 0.6I_{pink}(X) + 0.7_{yellow}(X)$$

$$\mu_{A_2}(X) = 0.7I_{black}(X) + 0.2I_{red}(X) + I_{brown}(X) + 0.4_{pink}(X) + 0.3_{yellow}(X)$$

$$\mu_{A_3}(X) = 0.3I_{black}(X) + 0.1I_{red}(X) + 0.5_{pink}(X) + 0.3_{yellow}(X)$$

7.2 Fuzzy Sample Mean for Fuzzy Data

Definition 7.2 Fuzzy sample mean (data with multiple values)

Let U be the universal set (a discussion domain), $L = \{L_1, L_2, \cdots, L_k\}$ be a set of k-linguistic variables on U, and $\{Fx_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \ldots + \frac{m_{ik}}{L_k}, i = 1,2,\ldots,n\}$ be a sequence of random fuzzy *sample on* U , $m_{ij}(\sum_{j=1}^{k} m_{ij}=1)$ *is the memberships with respect to* L_j *.Then, the Fuzzy sample mean is defined as*

$$F\overline{x}=\frac{\frac{1}{n}\sum_{i=1}^{n} m_{i1}}{L_1}+\frac{\frac{1}{n}\sum_{i=1}^{n} m_{i2}}{L_{i2}}+\ldots+\frac{\frac{1}{n}\sum_{i=1}^{n} m_{ik}}{L_k}.$$

Definition 7.3 Fuzzy sample mean (data with interval values)

Let U *be the universe set, and* $\{Fx_i=[a_i,b_i],a_i,b_i\in R,i=1,\ldots,n\}$ *be a sequence of random fuzzy sample on* U. *Then the fuzzy sample mean value is defined as* $F\overline{x}=[\frac{1}{n}\sum_{i=1}^{n} a_i,\frac{1}{n}\sum_{i=1}^{n} b_i]$.

Example 7.3 Let the x_1 =[2,3], x_2 = [3,4], x_3 = [4,6], x_4 = [5,8], x_5 =[3,7] be the beginning salary for 5 new master graduated students. Then fuzzy sample mean for the beginning salary of the graduated students will be

$$F\overline{x}=[\frac{2+3+4+5+3}{5},\frac{3+4+6+8+7}{5}]=[3.4,\ 5.6]$$

Example 7.4 Let the time series $\{x_t\}=\{0.8$，1.6，2.8，4.2，3.6，3.1，4.3，$3.5\}$，be the variation of a stock's value with 8 days . We can see the total range is 4.3 - 0.8＝3.5. We give a equal interval partition for [0.8, 4.3], say U＝{(0,1)，(1,2)，(2,3)，(3,4)，(4,5)}. The linguistic variable with respect the intervals are：very low $=L_1\propto(0,1]$，low $=L_2\propto(1,2]$，medium $=L_3\propto(2,3]$，high $=L_4\propto(2,3]$, very high $=L_5\propto(2,3]$; where $\propto$ means respect to. The medium for each partition are $\{m_1=0.5$，$m_2=1.5$，$m_3=2.5$，$m_4=3.5$，$m_5=4.5\}$. Since x_1 falls between 0.5

and 1.5. $\frac{1.5-0.8}{1.5-0.5}=0.7\in L_1$, $\frac{0.8-0.5}{1.5-0.5}=0.3\in L_2$ the fuzzy value of x_1 is $F_1=(0.7,0.3,0,0,0)$. Similarly, we get the other fuzzy samples as follows

	Very low=1	Low=2	Medium=3	High=4	Very high=5
F_1	0.7	0.3	0	0	0
F_2	0	0.9	0.1	0	0
F_3	0	0	0.7	0.3	0
F_4	0	0	0	0.3	0.7
F_5	0	0	0	0.9	0.1
F_6	0	0	0.4	0.6	0
F_7	0	0	0	0.2	0.8

The fuzzy expected value for the time series is

$$F\overline{x}=\frac{0.7/8}{1}+\frac{(0.3+0.9)/8}{2}+\frac{(0.1+0.7+0.4)/8}{3}$$

$$+\frac{(0.3+0.3+0.9+0.6+0.2+0.1)}{4}+\frac{(0.7+0.8+0.1)}{5}$$

$$=\frac{0.09}{1}+\frac{0.15}{2}+\frac{0.15}{3}+\frac{0.41}{4}+\frac{0.2}{5}$$

7.3 Fuzz Sample Median for Fuzzy Data

Definition 7.4 Fuzzy sample median (data with multiple values)

Let U be the universe set (a discussion domain), $L=\{L_1,L_2,\cdots,L_k\}$ be a ordered set of k-linguistic variables on U, and $\{x_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+\ldots+\frac{m_{ik}}{L_k}, i=1,2,\ldots,n\}$ be a sequence of random fuzzy *sample on* U , *Let* $S_j = \sum_{i=1}^{n} m_{ij}, j = 1,2,\cdots,k$, $T = 1\cdot S_1 + 2\cdot S_2 + \ldots + k\cdot S_k$. *Then, the minimum* L_j *such that* $\sum_{i=1}^{j} S_i \geq [\frac{T}{2}]$ ($[\frac{T}{2}]$ *means the largest integral that less or equal than* $\frac{T}{2}$) *is called the Fuzzy median of* $\{x_i\}$ *this sample. That is*

$$Fmedian(x_i) = \{ L_j : minimum\ j\ such\ that\ \sum_{i=1}^{j} S_i \geq [\frac{T}{2}] \} \tag{7.5}$$

Definition 7.5 Fuzzy sample median (data with interval values)

Let U *be the universe set, and* $\{Fx_i = [a_i, b_i], a_i, b_i \in R, i = 1,\ldots,n\}$ *be a sequence of random fuzzy sample on* U . *Let* c_j *be center of the interval of* $[a_i, b_i]$ *and* l_j *be the length of* $[a_i, b_i]$. *Then the fuzzy sample median is defined as:*

$$Fmedian = (c;r)\ ,\ c = median\,\{c_j\}, r = \frac{median\{l_i\}}{2} \tag{7.6}$$

Note that for the ordered intervals/terms, the fuzzy median, like the traditional median computing, is the midway of the cumulative memberships, which links the human thought to a more continuous flow rather than as discrete.

The following statements aim to illustrate certain heuristic properties applications for the presented definitions of fuzzy median and discuss some valuable factors.

Example 7.5 Suppose a community wants to set up a penalty for an environmental pollution. Hence they invite 5 experts to judge the penalty. Table 4.1 shows amount of penalty that the 5 experts propose.

Table 7.3 The amount of environmental penalty for 5 experts

Five experts	*A*	*B*	*C*	*D*	*E*
Environmental penalty	*3~5*	*6~8*	*5~8*	*7~10*	*13~19*

Since the penalty proposed by the expert E is much too high than the others did, if we use the average (mean) to decide the amount of penalty, it will be unfair since the result will be dominated by the E expert. However, using the fuzzy median seems an appropriate idea when the traditional median does not work. The computational process for the fuzzy median is shown below:

1. Find center points for each estimated interval, which are 4, 7, 6.5, 8.5, and 16. Hence the median is 7.
2. Find the width of each estimated interval, which are 2, 2, 3, 4, and 6. Hence the median is 3 (radius = 1.5)
3. According to Definition 3.4, the fuzzy median is [5.5, 8.5]

7.4 Fuzzy Sample Mode

Traditional statistics deals single answer or certain range of the answer through sampling survey, and unable to sufficiently reflect the complex thought of an individual. If people can use the membership function to express the degree of their feelings based on their own choices, the answer presented will be closer to real human thinking. Therefore, to collect the information based on the fuzzy mode should be the first step to take. Since a lot of times, the information itself embedded with uncertainty and ambiguity. It is nature for us to propose the fuzzy statistics, such as fuzzy mode and fuzzy median, to fit the modern requirement. In this and next section we demonstrate the definitions for fuzzy mode and fuzzy median generalized from the traditional statistics. The discrete case is simpler than the continuous one's.

Definition 7.6 Fuzzy sample mode (data with multiple values)

Let U be the universal set (a discussion domain), $L=\{L_1, L_2, \cdots, L_k\}$ a set of k-linguistic variables on U, and $\{FS_i, i=1,2\cdots,n\}$ a sequence of random fuzzy sample on U. For each sample FS_i, assign a linguistic variable L_j a normalized membership

$m_{ij}(\sum_{j=1}^{k} m_{ij}=1)$, let $S_j=\sum_{i=1}^{n} m_{ij}$, $j=1,2,\cdots,k$. Then, the maximum value of S_j (with respect to L_j) is called the fuzzy mode (FM) of this sample. That is $FM=\{L_j \mid S_j=\max_{1\le i\le k} S_i\}$.

Note : A significant level α for fuzzy mode can be defined as follows: Let U be the universal set (a discussion domain), $L=\{L_1,L_2,\cdots,L_k\}$ a set of k-linguistic variables on U, and $\{FS_i, i=1,2\cdots,n\}$ a sequence of random fuzzy sample on U. For each sample FS_i, assign a linguistic variable L_j a normalized membership $m_{ij}(\sum_{j=1}^{k} m_{ij}=1)$, let $S_j=\sum_{i=1}^{n} I_{ij}$, $j=1,2,\cdots,k$ $I_{ij}=1$ *if* $m_{ij}\ge\alpha$, $I_{ij}=0$ *if* $m_{ij}<\alpha$, α is the significant level. Then, the maximum value of S_j (with respect to L_j) is called the fuzzy mode (FM) of this sample. That is $FM=\{L_j \mid S_j=\max_{1\le i\le k} S_i\}$. If there are more than two sets of L_j that reach the conditions, we say that the fuzzy sample has multiple common agreement.

Definition 7.7 Fuzzy sample mode (data with interval values)

Let U be the universal set (a discussion domain), $L=\{L_1,L_2,\cdots,L_k\}$ a set of k-linguistic variables on U, and $\{FS_i=[a_i,b_i],$ $a_i,b_i\in R, i=1,2,\cdots,n\}$ be a sequence of random fuzzy sample on U. For each sample FS_i, if there is an interval $[c, d]$ which is covered by certain samples, we call these samples as a cluster. Let MS be the set of clusters which contains the maximum number of sample, then the fuzzy mode FM is defined as

$$FM=[a,b]=\{\cap[a_i,b_i] \mid [a_i,b_i]\subset MS\}.$$

If $[a,b]$ does not exist (i.e. $[a,b]$ is an empty set), we say this fuzzy sample does not have fuzzy mode.

Example 7.6 Suppose eight voters are asked to choose a chairman from four candidates. Table 7.4 is the result from the votes with two different types of voting: traditional response versus fuzzy response.

Table 7.4 Response comparison for the eight voters

Candidate / Voter	*traditional response*				*fuzzy response*			
	A	B	C	D	A	B	C	D
1		∨				0.7	0.3	
2	∨				0.5		0.4	0.1
3				∨			0.3	0.7
4			∨		0.4		0.6	
5		∨				0.6	0.4	
6				∨	0.4		0.4	0.6
7		∨				0.8	0.2	
8			∨				0.8	0.2
Total	1	3	2	2	1.3	2.1	3.5	1.6

From the traditional voting, we can find that three are three person vote for B. Hence the mode of the vote is B. However, from the fuzzy voting, B only gets a total membership of 2.1, while C gets 3.4. Based on traditional voting, B is elected the chairperson, while based on the fuzzy voting or membership voting, C is the chairperson. The voters' preference isreflected more accurately in fuzzy voting, C deserves to be the chairperson more than B does.

7.5 Heuristic Properties Related to the Fuzzy Statistics

Though features of a population can be illustrated by certain statistical parameters, there we still many characteristics which are left out, such as expectation, medium, and mode. Especially, when we are going to investigate the public opinions in the social science research, traditional parameters don't seem to be enough for application. In order to depict the whole picture more carefully, we implement the fuzzy mode.

Let U be the universe set, $L=\{L_1, L_2, \cdots, L_k\}$ be a set of k-linguistic variables on U , and $\{FS_i, i=1,2,\cdots,n\}$ be a sequence of random fuzzy sample on U (data with discrete fuzzy number). The following properties aim to illustrate the useful applications for the presented definition of fuzzy mode and discuss some valuable properties. We will also compare these two types of modes with the traditional ones.

Property 7.1

No matter the fuzzy samples come from continuous type or discrete type, the fuzzy sample median and fuzzy sample mean always exist and is unique, but the fuzzy sample mode may not exist.

Property 7.2

For discrete fuzzy sample, if the center of sample interval and sample interval come from an symmetric membership functions, then the center of fuzzy sample mean will equal to the center of fuzzy sample median.

Property 7.3

For continuous fuzzy sample, if the underlying distribution comes from the right (left) distribution, then the interval of fuzzy sample value will be greater (less) than that of fuzzy sample median.

Property 7.4

For discrete fuzzy samples, we can use the significant level α to adjust the appropriate sample size n.

Property 7.5

If the maximum membership m_{ij} in each fuzzy sample Fx_i is larger than the significant level α and located at L_j, then the fuzzy sample mode is consistent with the traditional mode.

Property 7.6

If there exists a membership m_{ij} in each fuzzy sample Fx_i with values $m_{ij} > 0.5$, then fuzzy mode consists with the traditional mode for any significant level $\alpha \geq 0.5$.

Property 7.7

If there are samples whose maximum membership falls on two or more of $L=\{L_1, L_2, \cdots, L_k\}$ *values. Then we can not compute the traditional mode without discarding these samples. However, by choosing an appropriate significant level* α*, we can compute it by*

$$S_j = \sum_{i=1}^{n} I_{ij} \text{ and } FM = \{L_j \mid S_j = \max_{1\le i\le k} S_i\} \text{ to get the fuzzy mode.}$$

Property 7.8

For samples with the continuous type, if the sample comes from a symmetric and unicorn distribution, then the center of its fuzzy sample mode will be coincided with the center of the Fuzzy sample median.

Property 7.9

For samples with the continuous type, we can use the significant level α *to control the number of samples which the fuzzy sample median inherits by itself.*

Note: If the significant level α is chosen to be too large, the value of fuzzy mode will be low. If we lower the significant level α, the value of fuzzy mode will increase. So, the degree of significant level α is an important point that will influence the state of fuzzy mode. The prior experience can help us to choose an appropriate significant level α according to the human thought or social utility.

7.6 Miscellaneous Applications

Example 7. 7 Evaluate the pollution problem (fuzzy mode application: data with multiple values)

In a field study, we ask 12 experts to evaluate the degree of pollution P ($0< P<1$) for a special topic. The Language variables L={ L_1 =OK, L_2 = Moderate, L_3 = Unhealthful, L_4 = Very unhealthful, L_5 = Hazardous. Table 5.1 and 5.2 illustrate the results from the 12 experts.

Table 7.5 Memberships for the 12 experts of environmental protection

Pollution / *Experts*	*o.k*	*moderate*	*unhealthful*	*very unhealthful*	*hazardous*
1	0.5	0.4	0.1		
2		0.4		0.6	
3		0.4	0.6		
4		0.6	0.4		
5	0.4	0.6			
6				0.3	0.7
7		0.4		0.6	
8			0.4	0.6	
9				0.2	0.8
10		0.1	0.3	0.6	
11	0.7	0.3			
12		0.7	0.3		
Total	1.6	3.9	2.1	2.9	1.5

Table 7.6 Pollution evaluation (traditional response)

Pollution / *Expert*	*o k*	*moderate*	*unhealthful*	*very unhealthful*	*hazardous*
1	∨				
2				∨	
3		∨			
4			∨		
5		∨			
6					∨
7				∨	
8				∨	
9					∨
10				∨	
11	∨				
12		∨			
Total	2	3	1	4	2

From Table 5.2, we find that there are four experts choose *very unhealthful*. Hence the mode is *very unhealthful*. But if we examine the Table 7.1, we can find that the item *very unhealthful* gets the total memberships 2.9, which is much less than the maximum membership *unhealthful*=3.9 of this survey. Comparing the traditional result at Table 5.2 with that of Table 5.1, it is clear that the fuzzy mode enables to be more robust in recording the data. In other word, fuzzy mode can more realistically reflect the common agreement than the traditional mode does.

Example 7.8 7-11 chain stores plan to introduce lunch box services (fuzzy mode application: data with interval values)

In order to understand consumers' pattern, the manager decides to take a survey to investigate the price of a lunch box which will be taken for granted and acceptable by consumers. They randomly select 50 customers reside in Taipei city. The investigator ask them, after they ate the lunch boxes, to fill up the questionnaire which price with a list single value and an interval value, which represents an acceptable range.

Table 7.8 Prices of lunch boxes (NT dollars)

Fuzzy answer	*[35,45]*		*[40,45]*		*[49,55]*		*[49,59]*		*[55,59]*		*[59,65]*		*[45,59]*		
Frequency	*5*		*2*		*15*		*10*		*8*		*4*		*6*		
Traditional answer	*35*	*45*	*40*	*45*	*49*	*55*	*49*	*55*	*55*	*59*	*59*	*65*	*45*	*55*	*59*
Frequency	*3*	*2*	*1*	*1*	*10*	*5*	*6*	*4*	*2*	*8*	*3*	*1*	*2*	*3*	*1*

From the above result, we could obtain the fuzzy mode is between NT$*49* to $*55* with size *31*. While based on the traditional answer, it is easy to find that the mode is NT$*49*, which is an accepted price for 16 persons. Note that there are still *14* customers choosing NT$55 as a reasonable price of a lunch box. Therefore we had better conclude that it is more realistic to say that the reasonable price of a lunch box will be $*49* to $*55*.

Example 7.9 The Design for English Lecture Hours in Primary School (fuzzy sample median application: interval data)

The design for lecture hours of different subjects in primary school should be carefully discussed by teachers and scholars. Assume a researcher would like to collect the information about rational lecture hours of English for primary school. So, we invite 30 teachers and scholars to set the rational lecture hours so that the result can be applied into nine-grade whole curriculum. The question is, "In your opinion, what will be the optimal English lecture hours per week in primary school?"

We ask people to fill out the two types of surveys, for instance, write range in the fuzzy survey and mark certain lecture hours in the traditional survey. So, every sample will contain both the results from the fuzzy survey and traditional survey. Table 5.4 shows that, based on the traditional survey, the mode is only *9* which means *9* out of *30* (less than 1/3) people would like *6* English lecture hours per week. However according to the definition of fuzzy mode, Table 5.4 also shows that up to *19* of *30* (almost 2/3) people would like (*generally agree*) *5~7* lecture hours per week based on the fuzzy survey.

Table 7.9 The English Lecture Per Week for Primary Schools

<table>
<tr><td>Fuzzy Hours</td><td>3~6</td><td colspan="2">4~7</td><td colspan="2">5~7</td><td colspan="2">6~9</td><td colspan="3">3~8</td><td colspan="2">5~10</td><td colspan="2">7~10</td></tr>
<tr><td>Number of People</td><td>2</td><td colspan="2">5</td><td colspan="2">6</td><td colspan="2">3</td><td colspan="3">5</td><td colspan="2">3</td><td colspan="2">6</td></tr>
<tr><td>Exact Hours</td><td>4</td><td>4</td><td>5</td><td>5</td><td>6</td><td>6</td><td>7</td><td>3</td><td>5</td><td>6</td><td>5</td><td>8</td><td>7</td><td>8</td></tr>
<tr><td>Number of People</td><td>2</td><td>4</td><td>1</td><td>2</td><td>4</td><td>2</td><td>1</td><td>1</td><td>1</td><td>3</td><td>2</td><td>1</td><td>4</td><td>2</td></tr>
</table>

Example 7.10 How many people should be hired? (fuzzy sample median application: interval data with α-cut)

A company wants to establish a new office. So the boss asks for 5 related managers to evaluate the number of newcomers that should be hired. Table 5.5 shows the *5* managers' evaluation as well as the membership function.

Table 7.10 Five managers' evaluation as well as the membership functions.

Numbers / Manager	1~3	3~6	6~9	9~12	12~15	15~18	18~21	21~24
A	0.4	0.6	0	0	0	0	0	0
B	0	0	0.3	0.7	0	0	0	0
C	0	0	0	0	0.7	0.3	0	0
D	0	0	0	0	0	0.1	0.2	0.7
E	0	0.4	0.6	0	0	0	0	0
Total	0.4	1.0	0.9	0.7	0.7	0.4	0.2	0.7

The significance of Table 5.6 lies on the fact that there is little overlaps among *5* experts' answers. This is what the α -cut can jump in and resolve the problem without bias. Table 5.8 shows the sum memberships via different significant level α=0.4 and α=0.7. It is clear to find that under α=*0.4*, the fuzzy median for the answer "*How many people should be hired* is *6~9* people". While under α=*0.7*, the fuzzy median for the same question is *9~12* people.

Table 7.11 Fuzzy median for 5 managers' evaluation

Numbers / Sum	1~3	3~6	6~9	9~12	12~15	15~18	18~21	21~24
Membership	0.4	1	0.9	$\boxed{0.7}$	0.7	0.4	0.2	0.7
Membership(= 0.4)	0.4	1	$\boxed{0.9}$	0.7	0.7	0	0	0.7
Membership(= 0.7)	0	1	0.9	$\boxed{0.7}$	0.7	0	0	0.7

Example 7.11 Penalty evaluation for a river pollution (fuzzy median/mean application: linguistic data with weight)

Suppose a community wants to set up the penalty for a river pollution. They ask *10* experts to estimate the total amount of the penalty. Table 7.12 is the result of the surveys.

Table 7.12 Penalty for a river pollution

Factors / *Expert*	*Water* w_1 *=0.4*	*solid waste* w_1 *=0.25*	*Odor* w_1 *=0.2*	*Chemical toxins* w_1 *=.1*	*Noise* w_1 *=0.05*
1	[10,12]	[15,20]	[12,15]	[22,30]	[8,10]
2	[8, 9]	[8,10]	[10,13]	[6, 8]	[12,13]
3	[5, 7]	[20,25]	[19,20]	[9,11]	[20,22]
4	[12,15]	[20,28]	[11,12]	[20,23]	[15,18]
5	[20,30]	[14,19]	[10,20]	[14,16]	[16,20]
6	[7, 9]	[16,21]	[13,15]	[8, 9]	[25,30]
7	[8,10]	[10,12]	[22,24]	[10,11]	[31,35]
8	[5, 6]	[7, 8]	[17,19]	[12,15]	[20,23]
9	[11,12]	[11,14]	[22,28]	[18,20]	[10,12]
10	[11,13]	[22,26]	[14,16]	[10,12]	[16,22]
Fuzzy Median	[9,11]	[14.5, 19.5]	[14,16]	[11,13]	[17,20]
Fuzzy mean	[9.7,12.3]	[14.3,18.3]	[15,18.2]	[12.9,15.5]	[17.3,20.5]

unit=million dollars

The fuzzy median estimation process is as follows:

(1) Deciding the influence factors about the pollution.

(2) Applying the fuzzy ordering method to calculate the weights $\{w_1, w_2, ..., w_k\}$.

(3) Calculating the sample fuzzy median *FM* of the penalty for each item.

(4) Calculating the total penalty via soft computing *FM** $\{w_1, w_2, ..., w_k\}$.

Take the first term quality of water as an example: (1)the center point and the radius [*m;r*] are [10;1], [17;2.5], [15;1], [12;1], [18.5,1.5]. From Definition 3.4, we can find that the (10,1) is the fuzzy median of these 10 data, that is [9,11] for item "*water*". Finally multiply the weight by the center of these fuzzy medians, we can get the amount of penalty for a river pollution, which is

$$10 \cdot 0.4 + 17 \cdot 0.25 + 15 \cdot 0.2 + 12 \cdot 0.1 + 18.5 \cdot 0.05 = 13.4$$

While multiply the weight by the center of these fuzzy means, we can get the amount of penalty for a river pollution becomes

$$11 \cdot 0.4 + 16.3 \cdot 0.25 + 16.6 \cdot 0.2 + 19.2 \cdot 0.1 + 19.4 \cdot 0.05 = 14.8.$$

7.7 Concluding Remarks

Fuzzy statistical analysis grows as a new discipline from the necessity to deal with vague samples and imprecise information caused by human thought in certain experimental environments. In this chapter, we made an attempt to link the gap between the binary logic based on multiple choice survey with a more complicated yet precise fuzzy membership function assessment, such as fuzzy mode, fuzzy median and fuzzy mean, fuzzy weight and α-cut etc. We carefully revealed how theses factors can be properly and easily utilized in various fields to reveal the contradictory characteristics of human concepts. Through these processes, human ideas are no longer presented as discrete bust as a natural and continuous flow. There are illustrated examples demonstrated to explain how to find the fuzzy mode and fuzzy median, and how to use the results to help people reaching their decisions.

However, there are still some problems we need to investigate in the future:

1. We can further the research on data simulation so that we can understand features of the fuzzy linguistic, multi-facet assessment, and the balance of the moving consensus. Moreover, the choice of different significant α-cut will influence the statistical result. An appropriate criterion for selecting significant α-cut should be investigated in order to reach the best common agreement of human beings.

2. There are other types of membership functions we could explore in the future. For the fuzzy mode of continuous type, we can extend the uniform and triangular types of membership functions to non-symmetric or multiple peaks types.
3. The degree of correlation between the ambiguity of human thoughts and the uncertainty of human behavior. That is, how true does the human behavior honestly reflect the human thought.

Chapter 8 Tests of Hypothesis: Means

In many expositions of fuzzy methods, fuzzy techniques are described as an alternative to a more traditional statistical approach. In this chapter, we present a class of fuzzy statistical decision processes in which hypothesis testing can be naturally reformulated in terms of interval-valued statistics. We provide the definitions of fuzzy mean, fuzzy distance as well as investigating their related properties. We also give some empirical examples to illustrate the techniques and to analyze fuzzy data. Empirical studies show that fuzzy hypothesis testing with soft computing for interval data are more realistic and reasonable in the social science research.

8.1 Introduction

A statistical test is general conducted by means of a hypothesis testing for which the probability distribution is determined by the assumption that the null hypothesis H_0 is true. Under the significant level α, a critical region is computed such that if the observed statistics falls in the critical region, we reject the null hypothesis. The statistical test provides information from which we can decide the significance of the increase (or decrease) in any experimental result.

However there are some vague information, formulated by terms from natural language, which is not easy to describe in statistical terms. To handle this information and knowledge, it is natural to use intelligent computing techniques. Fortunately, in many expositions of intelligent computing methods, fuzzy techniques are described as an alternative to a more traditional statistical approach. Such a description makes fuzzy techniques difficult to understand and difficult to accept for researchers who are accustomed to statistical methods.

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 129–144 (2006)
www.springerlink.com

In this chapter we will introduce the process of hypothesis testing of fuzzy statistics via discrete fuzzy sample and continuous fuzzy sample. The definitions of fuzzy mean for two kinds of fuzzy data are proposed at Chapter 7. Using these definitions we are able to set up the fuzzy testing hypothesis such as fuzzy equals to and the fuzzy belongs to.

Although the statistical methods used to testing fuzzy sample mean are essentially based on the traditional decision theory, extending Neyman-Peason's lemma about most powerful test they have not been aggressively not austerely investigated for two reasons: (i) soft-computing for critical regions with fuzzy numbers is still not identified. (ii) distributions for fuzzy populations are vague, incomplete or unknown.

Since our main objectives is to promote the understanding of these two classes of techniques - statistics and fuzzy - to researchers who may only know well one of theses techniques, we go into some detail explaining the basics of techniques. A reader who is well familiar with one or both of these techniques is advised to at least browse through the basics.

In reality, more complex situations are possible, in which an expert is not 100% sure whether a given estimate x_i is possible; in this case, we can no longer use polling to get numerical characteristics of the expert knowledge. Fuzzy set methodology can handle such more complex situation as well. The above polling method of eliciting the values of the membership functions is only one of the many known elicitation techniques; see, e.g.,[52], [87],[109], [110].

In this paper, we present a class of fuzzy statistical decision processes in which hypothesis testing can be naturally reformulated in terms of interval-valued statistics. To describe these situations we will start with a brief motivation of traditional statistical techniques, and then give a brief motivation of the corresponding fuzzy methods, and then describes the relation between these two classes of techniques. We provide the definitions of fuzzy mean, fuzzy distance as well as investigating of their related properties. We also give some empirical examples to illustrate the techniques and to analyze fuzzy data. Empirical studies show that fuzzy hypothesis testing with soft computing for interval data are more realistic and reasonable in the social science research. Finally certain comments are suggested for the further studies.

Our result is in good agreement with a general result from [111], according to which an arbitrary fuzzy set can be interpreted in statistical terms: namely, as a random *set*. For the latest developments in this area, see, e.g., [112]. We hope that this reformation will make the corresponding fuzzy techniques more acceptable to researchers whose only experience is in using traditional statistical methods.

Definition 8.1 Defuzzification for discrete fuzzy data

Let D be a fuzzy sample on universe domain U with ordered linguistic variable $\{ L_i; i = 1,\dots,k \}$. $\mu_D(L_i) = m_i$ *is the membership with respect to* L_i, $\sum_{i=1}^{n} \mu_D(L_i) = 1$. *We call* $D_f = \sum_{i=1}^{k} m_i L_i$ *the defuzzification value for the discrete fuzzy sample D.*

Example 8.1 Let $D = \frac{0}{1} + \frac{0.2}{2} + \frac{0.5}{3} + \frac{0.3}{4} + \frac{0}{5}$ be a discrete fuzzy sample on the ordered *linguistic variable* $\{ L_1 = 1, L_2 = 2, L_3 = 3, L_4 = 4, L_5 = 5 \}$. Then the defuzification value for the fuzzy sample *D is*

$$D_f = \sum_{i=1}^{k} m_i L_i = 0 \cdot 1 + 0.2 \cdot 2 + 0.5 \cdot 3 + 0.3 \cdot 4 + 0 \cdot 5 = 3.1$$

Definition 8.2 Defuzzification for interval fuzzy data

Let C be a fuzzy sample on universe domain U with support $[a,b]$, *that is, C is a fuzzy interval of* $\mathcal{R}$ *with support [a,b]. Then* $C_f = \frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx}$ *is called the defuzzification value of C.*

Example 8.2 Let C be a fuzzy sample with support [0,2] and membership function 1/2. Then $C_f = \dfrac{\int_a^b xf(x)dx}{\int_a^b f(x)dx} = \dfrac{\int_0^2 (\frac{x}{2})dx}{\int_0^2 (\frac{1}{2})dx} = 1$ is called the defuzzification value of *C*.

8.2 Soft Distance with Fuzzy Samples

In order to set up an appropriate testing hypothesis on fuzzy data, it is necessary for us to give definitions about measurement of distance of fuzzy set. In the following, we set up firstly, the definition of fuzzy distance with fuzzy interval data. The definition is difference from the traditional interval operations. Our consideration is concentrated on the statistical point of view.

Definition 8.3 Distance for discrete fuzzy sample

Let A, B be two fuzzy samples on the universe set U with membership functions μ_A, μ_B. *Then*

(a) Euclidian distance,

$$d_E(A,B) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(\mu_A(x_i) - \mu_B(x_i))^2} \quad (n = |U| \text{ is the cardinality of } U)$$

(b) Hamming distance: (the number of mismatched component of A and B)

$$d_H(A,B) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i : \mu_A(x_i) \neq \mu_B(x_i)\}}(x_i) \quad (|U| \text{ is the cardinality of U})$$

Definition 8.4 Distance of interval fuzzy samples

Let A, B be two fuzzy data with membership functions $\mu_A(x) = f(x), \text{if } x \in [a,b]$, $0 \le f(x) \le 1$, $\mu_B(y) = g(y)$, $y \in$ [c,d], $0 \le g(y) \le 1$. *We give three definitions of distance:*

d_1 *(A,B)=inf*$\{|x = y| : x \in A, y \in B\}$,

d_2 *(A,B)=sup* $\{|x = y|: x \in A, y \in B\}$,

d_3 *(A,B)=inf* $\{\varepsilon_1, \varepsilon_2\}$, d_4 *(A,B)=sup* $\{\varepsilon_1, \varepsilon_2\}$;

d_4 *(A,B)=sup* $\{\varepsilon_1, \varepsilon_2\}$

where $\varepsilon_1 = \inf\{\varepsilon : [c,d] \subset [a-\varepsilon, b+\varepsilon]\}$, $\varepsilon_2 = \{\varepsilon : [a,b] \subset [c-\varepsilon, d+\varepsilon]\}$.

Example 8.3 *Let A, B be two fuzzy data with support* [1,3], [2,5]. *Then*

d_1 *(A,B)=inf* $\{|x = y|: x \in A, y \in B\}$ *=0*

d_2 *(A,B)=sup* $\{|x = y|: x \in A, y \in B\}$ *=4*

$\varepsilon_1 = \inf\{\varepsilon : [c,d] \subset [a-\varepsilon, b+\varepsilon]\}$ *=2,* $\varepsilon_2 = \{\varepsilon : [a,b] \subset [c-\varepsilon, d+\varepsilon]\}$ *=1*

d_3 *(A,B)=inf* $\{\varepsilon_1, \varepsilon_2\}$ *=1,*

d_4 *(A,B)=sup* $\{\varepsilon_1, \varepsilon_2\}$ *=2*

It is a new research topic about the hypothesis testing of fuzzy mean with interval values. First of all, we will give a brief definition about the defuzzification. Then under the fuzzy significant level δ, we make a one side or two side testing. These methods are a little different from traditional significant level α. In order to get the robustic characteristic properties, we will set up the rejection area level F_δ, according to the fuzzy population.

Definition 8.5 Fuzzy equality for discrete data

Let U be a universe domain, $L = \{L_1, L_2, \cdots, L_k\}$ *be sequence of rank ordering of linguistic variables on U,* $\{ X_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \ldots + \frac{m_{ik}}{L_k},\ , i = 1,2\}$ $\sum_{j=1}^{k} m_{ij} = 1$ *are two random samples from U. If* $m_{1j} = m_{2j}$ *; j =1,2,...k. Then we say that sample* X_1 *fuzzily equals to* X_2, *denoted by* $X_1 \approx_F X_2$.

The fuzzy index number for discrete type is as follows

Definition 8.6 Fuzzy index for equality of discrete data

Let U be a universe domain, $L=\{L_1,L_2,\cdots,L_k\}$ *be sequence of rank ordering of linguistic variables on U,* $\{X_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+...+\frac{m_{ik}}{L_k},\ ,i=1,2\}$ $\sum_{j=1}^{k}m_{ij}=1$ *are two random samples from U. The defuzzification value for the discrete fuzzy sample* X_i *is* $D_{if}=\sum_{j=1}^{k}m_{ij}L_j$ *. If* $D_{1f}=D_{2f}$*, then we say that* X_1 *fuzzy index equals to* X_2*, denoted by* $X_1\approx_I X_2$*,*

Definition 8.7 Fuzzy equality for interval data

Let A, B be two fuzzy data with membership functions $\mu_A(x)=f(x), if\ x\in[a,b]$, $0\le f(x)\le 1$, $\mu_B(y)=g(y),\ y\in[c,d]$, $0\le g(y)\le 1$. *If A, B have the same support and f, g are all convex functions then we say that A fuzzily equal to B, written as* $A=_F B$*, or briefly writed .* $A=_F[a,b]$.

For left unbounded or right unbounded, the definitions are similar.

Definition 8.8 Fuzzy belonging for interval data

Let A, B be two fuzzy data with membership functions $\mu_A(x)=f(x), if\ x\in[a,b]$, $0\le f(x)\le 1$, $\mu_B(y)=g(y),\ y\in$ [c,d], $0\le g(y)\le 1$. *If the support of A is contained in the support B, f, g are all convex functions, then we say A is fuzzy belongs to B, written as* $A\in_F B$*, or briefly writed .* $A\in_F[c,d]$.

For left unbounded or right unbounded, the definitions are similar.

8.3 Some Properties of Fuzzy Data

Property 8.1. *Let A, B be two fuzzy data with membership functions* $\mu_A(x)=f(x), if\ x\in[a,b]$, $0\le f(x)\le 1$, $\mu_B(y)=g(y),\ y\in$ [c,d], $0\le g(y)\le 1$. *The fuzzy equality implies fuzzy belonging. The converse is not true.*

Proof. If A$=_F$ [a,b],since [a,b]=[a,b] implies [a,b] $\subset$ [a,b]. Hence A$\in_F$[a,b].

Properties 8.2

(i)For any fuzzy set C with support [*m,n*] *and has no intersection with the support of A and B. If b < m, d < m and* $d_1(A,C)=d_1(B,C)$ *,* $d_2(A,C)=d_2(B,C)$. *Then* $A=_F B$

(ii) For any fuzzy set C with support [*m,n*] *and have no intersection with the support of A,B, if a > n, c > d and* $d_1(A,C)=d_1(B,C)$ *,* $d_2(A,C)=d_2(B,C)$ *then* $A=_F B$

Proof.

(i) If $d_1(A,C)=d_1(B,C)$, i.e. then *inf* $\{|x-z|: x\in A, y\in B\}$ = $\inf\{|y-z|: y\in B, z\in C\}$, then *m-b* = *m-d* we get *b*=*d*,
If $d_2(A,C)=d_2(B,C)$, i.e. sup $\{|x-y|: x\in A, y\in B\}$ = $\sup\{|y-z|: y\in B, z\in C\}$ then *n-a*=*n-c*, we get *a*=*c*,
hence $A=_F B$.

(ii) similar to (i)$\cdot$

Property 3.3 *(1)For any fuzzy set C with support* [*m,n*] *and has no intersection with the support of A and B. If* $A\in_F B$*, then* $d_1(A,C)\ge d_1(B,C)$ *and* $d_2(A,C)\le d_2(B,C)$. *The converse is not true*

Proof：We only prove *d* < *m* case. The other cases are similar.

Since $A\in_F B$, $a\ge c, b\le d$, $m\text{-}b\ge m\text{-}d$, and $n\text{-}a\le n\text{-}c$

So we have $d_1(A,C)\ge d_1(B,C)$ and $d_2(A,C)\le d_2(B,C)$.

Remark. On the other hand, if we choose [a,b]=[1,3], [c,d]=[15,18], [m,n]=[8,10] then $d_1(A,C)\ge d_1(B,C)$ and $d_2(A,C)\le d_2(B,C)$ but $A\notin_F B$.

Property 8.4. *Let* $[a,b] \cap [c,d] = \phi$ *and d-c* $\geq$ *b-a. Then* $d_3(A,B) = c-a$ *if* $b < c$ *and* $d_3(A,B) = b-d$ *, if* $a > d$.

Proof. Considering $b < c$, (the other cases are similar)

$\varepsilon_1 = \inf\{\varepsilon : [c,d] \subset [a-\varepsilon, b+\varepsilon]\}$ =*d-b*, $\varepsilon_2 = \inf\{\varepsilon : [a,b] \subset [c-\varepsilon, d+\varepsilon]\}$ = *c-a*. $\varepsilon_1 - \varepsilon_2$ = (d-c)-(b-a) $\geq$ 0, then $d_3(A,B) = \inf\{\varepsilon_1, \varepsilon_2\}$ =*c-a*

Property 8.5 *Let* $[a,b] \cap [c,d] \neq \phi$ *and* $[a,b] \not\subset [c,d]$ *and d-c* $\geq$ *b-a then* $d_3(A,B) = c-a$ *if* $a < c$, $d_3(A,B) = b-d$ *if* $b > d$

Proof. Since d-c $\geq$ b-a we have d-b $\geq$ c-a. The proof is similar to that of Property 8.4.

Property 8.6. *If* $[a,b] \subset [c,d]$ *then* $d_3(A,B) = 0$

Proof.

$$\varepsilon_1 = \inf\{\varepsilon : [c,d] \subset [a-\varepsilon, b+\varepsilon]\} > 0,$$

$$\varepsilon_2 = \{\varepsilon : [a,b] \subset [c-\varepsilon, d+\varepsilon]\} = 0, \text{ hence } d_3(A,B) = 0。$$

Property 8.7 *For any fuzzy set C with support [m,n] (i) if n-m* $\geq$ *d-c and* $A \in_F B$ *then* $d_3(A,C) \leq d_3(B,C)$ *(ii) if n-m* $\geq$ *d-c and* $A =_F B$ *then* $d_3(A,C) = d_3(B,C)$

Proof：(i) If $A \in_F B$ then $a \geq c$,$b \leq d$

case1：$d < m$, *then* $d_3(A,C) = m-a \leq m-c = d_3(B,C)$

case2：$c > n$, *then* $d_3(A,C) = b-n \leq d-n = d_3(B,C)$

case3：$c < m$, $a < m$, $d \geq m$, *then* $d_3(A,C) = m-a \leq m-c = d_3(B,C)$

case4：$c < m$, $a \geq m$, *then* $d_3(A,C) = 0 < m-c = d_3(B,C)$

case5 : $c \geq m$, $d \leq n$, then $d_3(A,C)=0=d_3(B,C)$

case6 : $c \leq n$, $b > n$, $d > n$, then $d_3(A,C)=b-n \leq d-n=d_3(B,C)$

case7 : $c \leq n, b \leq n, d > n$ then $d_3(A,C)=0 < d-n=d_3(B,C)$.

The proof is completed.

(ii) similar proof with (i).

8.4 Hypothesis Testing with Fuzzy Samples

It is a new research topic about the hypothesis testing of fuzzy mean with interval values. First of all, we will give a brief definition about the defuzzification. Then under the fuzzy significant level δ, we make a one side or two side testing. These methods are a little different from traditional significant level α. In order to get the robustic properties, we will set up the rejection area level F_δ, according to the fuzzy population.

Let $\overline{Fx}$ be the fuzzy sample mean , $\overline{x}_f$ be the *defuzzyfication* of $\overline{Fx}$. Under the fuzzy significant level F_δ, and the corresponding critical value F_δ, we want to testing H_0: $\overline{Fx}=F\mu_0$, where $F\mu_0$ is the fuzzy mean of the underlying population. Let μ is the defuzzification value of $F\mu$, then the above hypothesis becomes H_0: $\mu=\mu_0$.

Testing hypothesis of fuzzy (index) equality for discrete fuzzy means

1. *Let* $\{ x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \ldots + \frac{m_{ik}}{L_k}, \ , i=1,\ldots,n\}$, $\sum_{j=1}^{k} m_{ij} = 1$ *be a set of random samples from population U with fuzzy mean* $F\mu_0$. *Let* $\overline{x}_f$ *and* μ_0 *be the defuzzification values of* $\overline{Fx}$ *and* $F\mu_0$ *respectly..*

2. *Hypothesis:* H_0 : $F\mu = F\mu_0$ *vs.* H_1 : $F\mu \neq F\mu_0$

3. *Statistics: find* $\overline{Fx}$ *from a random sample* $\{x_i, i=1,\ldots,n\}$.

4. Decision rule: under the fuzzy significant level F_δ, *if* $|\overline{x}_f - \mu_0| > \delta$, *then reject* H_0

Note: for left side test $H_0 : F\mu \le F\mu_0$ vs. $H_1 : F\mu > F\mu_0$ under the fuzzy significant level F_δ, if $\mu_0 - \overline{x}_f > \delta$, we reject H_0. The right hand side testing is similar.

Hypothesis testing with continuous fuzzy means

1. Let Ω *be a universe domain with fuzzy mean* $[a,b]$, and $\{x_i = [x_{li}, x_{ui}], i = 1,\ldots,n\}$ *be a set of random sample*

2. Hypothesis: $H_0 : F\mu =_F [a,b]$ *vs.* $H_1 : F\mu \ne_F [a,b]$

3. Statistics: find $\overline{Fx} = [x_l, x_u]$ *from a random sample* $\{x_i, i = 1,\ldots,n\}$.

4. Decision rule: under the significant level F_δ, *find k=δr (where r=b-a)*, if $|x_l - a| > k$ or $|x_u - b| > k$ *then reject* H_0

Testing of fuzzy belonging with bounded samples

1. Hypothesis: $H_0 : F\mu \in_F$ [a,b] vs. $H_1 : F\mu \notin_F$ [a,b].

2. Statistics: find $\overline{FX} = [x_l, x_u]$ *from a random sample* $\{x_i, i = 1,\ldots,n\}$.

3. Decision rule: under the significant level F_δ, *find k=δr (where r=b-a)*,

if $x_l < a\text{-}k$ or $x_u > b+k$ *then reject* H_0

Testing of fuzzy belonging to with unbounded below samples

1. Hypothesis: $H_0 : F\mu \in_F F\mu_0 = (-\infty, b]$ vs. $H_1 : F\mu \notin_F F\mu_0$

2. Statistics: find $\overline{Fx} = (\infty, x_u]$ *from a random sample* $\{x_i, i = 1,\ldots,n\}$.

3. Decision rule: under the significant level F_δ, *find k=δr (where r is a constant)*, if $x_u > b+k$ *then reject* H_0

Testing of fuzzy belonging to with unbounded above sample

1. *Hypothesis:* H_0 : $F\mu \in_F \mu \in_F [a, \infty)$ vs. H_1 : $F\mu \notin_F [a, \infty)$

2. *Statistics: find* $\overline{Fx} = [x_l, \infty)$ *from a random sample* $\{x_i, i = 1,\ldots,n\}$.

3. *Decision rule: under the significant level* F_δ, *find k*=δ*r* (*where r is a constant*), if x_l < *a-k then reject* H_0

Example 8.4

A farmer wants to adapt a new cooking style of fried chicken from traditional techniques. He invites 5 experts to join the evaluating experiment. After they tested the new fried chicken, they are asked to give a fuzzy grading with: very unsatisfactory =1, unsatisfactory =2, no difference = 3, satisfactory = 4, very satisfactory=5. Table 8.1 shows the result of the 5 experts' evaluation

Table 8.1 Evaluation result for 5 experts

Expert	*1*	*2*	*3*	*4*	*5*
A	*0*	*0*	*0*	*0.7*	*0.3*
B	*0*	*0*	*0*	*0*	*1.0*
C	*0*	*0.4*	*0.6*	*0*	*0*
D	*0*	*0*	*0*	*0.8*	*0.2*
E	*0.1*	*0.9*	*0*	*0*	*0*

Let's set up the hypothesis testing for fuzzy index equal:

H_0 : $\overline{X}_f$ *=3 vs* H_1 : $\overline{X}_f \neq 3$.

Under the significant level $\delta = 0.1$, since $\overline{x}_f$ *=2.4* and μ_0 - $\overline{x}_f$ *=3-2.4= 0.6 > 0.1*. Hence we reject H_0. And since the sample fuzzy index $\overline{x}_f$ is less than 3, the manager will not apply this new cooking style.

Example 8.5

A company administrator wants to control the time of turning air-condition base on the energy saving reason. He feel that that the temperature over 28˚C will be hot and is the time to turn on. However, he wants to know how the other staff feeling. So, he asks for five staffs at random to investigate and then gets five data $[27,\infty)$, $[26,\infty)$, $[29,\infty)$, $[24,\infty)$, $[26,\infty)$.

Hypothesis H_0 ： $\mu=[28,\infty)$ *vs* H_1 ： $\mu\neq[28,\infty)$

After simple computation, we have $\overline{Fx}=[26.4,\infty)$. Under the significant level $\delta=0.2$, since $28\text{-}26.4>0.2$, we reject H_0 and suggest that turn on the air condition when the temperature is below 28˚C.

Example 8.6

The human resource department announced that 20 to 26 years old people request their salary between 20 thousands and 40 thousands with a deviation 5 thousands. The manger asks the statistics department to check it up. Suppose they find 10 young men between 20 and 26 years old, survey their request salary, the sample data are: [3,4], 1.8, [2,3], [4,6], [1.5,2] [3,4], 2, [2,3], [3,5], [2.5,4] (unit:10thousands).

Hypothesis H_0 ∶ $F\mu\in_F$ *[2,4] vs.* H_1 ∶ $F\mu\notin_F$ *[2,4]*.

We treat 1.8 as [1.8,1.8]. After simple computation, we get

$$\overline{FX}=\left[\frac{3+1.8+2+4+1.5+3+2+2+3+2.5}{10},\frac{4+1.8+3+6+2+4+2+3+5+4}{10}\right]$$

=[2.38,3.48]. Under the significant level δ=0.5, since 2.38＞2-0.5 and 3.48＜4+0.5 and 2.38＜2+0.5 but 3.48＜4-0.5,we do not accept what hnman resource departmemt says. i.e. We accept H_0 $\overline{Fx}\in_F[2,4]$.

8.5 Fuzzy χ^2-test of Homogeneity

Consider a K-cell multinomial vector $n=\{n_1,n_2,\ldots,n_k\}$ with $\sum_i n_i=n$. The *Pearson chi-squared test* ($\chi^2=\sum_i\sum_j\frac{n_{ij}-e_{ij}}{e_{ij}}$) is a well known statistical test for investigating the significance of the differences between observed data arranged in K classes and the theoretically expected frequencies in the K classes. It is clear that the large discrepancies between the observed data and expected cell counts will result in larger values of χ^2.

However, a somewhat ambiguous question is whether (quantitative) discrete data can be considered categorical and use the traditional χ^2 *-test*. For example, suppose a child is asked the following question: "*how much do you love your sister*?" If the responses is a fuzzy number (say, 70% of the time), it is certainly inappropriate to use the traditional χ^2 -test for the analysis. We will present a χ^2 -test for fuzzy data as follows:

Procedures for Testing hypothesis of homogeneity for discrete fuzzy samples

1. Sample : Let Ω *be a domain ,* $\{L_j, j=1,...,k\}$ *be ordered linguistic variables on* Ω *, and* $\{a_1, a_2, \cdots, a_m\}$ *and* $\{b_1, b_2, \cdots, b_n\}$ *are random fuzzy sample from population A, B with standerized membership function* mA_{ij}, mB_{ij}.

2. Hypothesis: Two populations A, B have the same distribution ratio. i.e

$$H_0 : F\mu_A =_F F\mu_B , \text{ where } F\mu_A = \frac{\frac{1}{m}MA_1}{L_1} + \frac{\frac{1}{m}MA_2}{L_2} + ... + \frac{\frac{1}{m}MA_k}{L_k}$$

$$F\mu_B = \frac{\frac{1}{n}MB_1}{L_1} + \frac{\frac{1}{n}MB_2}{L_2} + ... + \frac{\frac{1}{n}MB_k}{L_k} , \; MA_j = \sum_{i=1}^{m} mA_{ij} , \; MB_j = \sum_{i=1}^{n} mB_{ij} .$$

3. Statistics : $\chi^2 = \sum_{i\in A,B} \sum_{j=1}^{c} \frac{([Mi_j] - e_{ij})^2}{e_{ij}}$. *(In order to perform the Chi-square test for fuzzy data, we transfer the decimal fractions of* Mi_j *in each cell of fuzzy category into the integer* Mi_j *by counting 0.5 or higher fractions as 1 and discard the rest.*)

4. Decision rule：*under significance level α, if* $\chi^2 > \chi^2_\alpha(k-1)$*,then we reject* H_0.

Procedures for Testing hypothesis of homogeneity for interval fuzzy samples

1. Sample：Let Ω *be a discussion domain,* $\{L_j, j=1,\ldots,k\}$ *be ordered linguistic variables on the total range of* Ω*, and* $\{a_i=[a_{li}, a_{ui}], i=1,\ldots,m\}\}$ and $\{b_i=[b_{li}, b_{ui}], i=1,\ldots,n\}$ *and are random fuzzy sample from population A, B with standardized membership function* mA_{ij}, mB_{ij}.

2. Hypothesis: Two populations A, B have the same distribution ratio. i.e

$$H_0 : F\mu_A =_F F\mu_B, \quad \text{where } F\mu_A = \frac{\frac{1}{m}MA_1}{L_1} + \frac{\frac{1}{m}MA_2}{L_2} + \ldots + \frac{\frac{1}{m}MA_k}{L_k}$$

$$F\mu_B = \frac{\frac{1}{n}MB_1}{L_1} + \frac{\frac{1}{n}MB_2}{L_2} + \ldots + \frac{\frac{1}{n}MB_k}{L_k}, \quad MA_j = \sum_{i=1}^{m} mA_{ij},\ MB_j = \sum_{i=1}^{n} mB_{ij}.$$

3. Statistics： $\chi^2 = \sum_{i\in A,B}\sum_{j=1}^{c}\frac{([Mi_j]-e_{ij})^2}{e_{ij}}$. *(In order to perform the Chi-square test for fuzzy data, we transfer the decimal fractions of* Mi_j *in each cell of fuzzy category into the integer* Mi_j *by counting 0.5 or higher fractions as 1 and discard the rest.*)

4. Decision rule：*under significance level α, if* $\chi^2 > \chi^2_\alpha(k-1)$*,then we reject* H_0.

Example 8.7 DDP party wants to know the degree of support from an election. Suppose they are interested in how the sex will make a difference about the voting. They conduct a sampling survey and ask the people with two methods for reply: traditional reply and fuzzy reply. The result is as follows:

Table 8.2 Replies for peoples on the degree of party support

	Support of Parties			χ^2 -test of homogeneity	Support of parties			χ^2 -test of homogeneity
Category	DDP	KMT	others	χ^2 =8.27 > 5.99 $=\chi^2_{0.05}(2)$	DDP	KMT	others	χ^2 =3.78 < 5.99 $=\chi^2_{0.05}(2)$
Male	220	280	100		216.2	268.5	114.3	
Female	170	150	80		158.1	154.7	87.2	

Null Hypothesis: H_0 : there is no difference of the degree of support for parties. H_1 : there is no difference of the degree of support for parties. Under the significance level α =0.05, we can find that there exists difference Statistical testing conclusion: for traditional reply, we will reject the null hypothesis. While for the fuzzy reply, will accept the null hypothesis.

Example 8.6 In order to set up a sales strategy, the R&D of a supermarket manger want to know *the living expense(monthly)* between community *X and community Y.* They randomly choose 50 samples from *X and Y*. during the answering process, people are asked to write their living expense by interval instead of real number. For instance, they can write the living expense as: 1500~2500 with membership 0.7, 2500~4000 with membership 0.3. Then they sum up the memberships and get the following table 8.3

Table 8.3 Monthly living expense for community X and Y

	0~1500	1500~2500	2500~4000	4000~6000	6000+
X	2.8	10.3	19.7	14.2	5.0
Y	7.1	21.6	20.9	6.8	2.6

Null Hypothesis H_0 : The distribution (ratio) for living expense between is no difference. H_1 : *community X* has a higher living expense than *Y*.

Computing the statisitcs χ^2, we find $\chi^2=8.43>\chi^2_{0.05}(4)=7.78$. Hence under the significant level $\alpha=0.1$. We reject H_0 : The distribution (ratio) for living expense between is no difference. Examining again the data, we may say that the community *X* has a higher living expense than community *Y*.

Chapter 9 Fuzzy Time Series Analysis and Forecasting

The problem of system modeling and identification has attracted considerable attention during the past decades mostly because of a large number of applications in diverse fields. In this chapter we give a well defined fuzzy time series model with forecasting through the Markov fuzzy relation. An illustrative algorithm based on a fuzzy logic is developed to predicate the multivariate dynamic data.

On the other hand, the profit of investment does not lie solely in the accuracy of prediction, but in the degree of belief as well. The greater the degree of belief is, the more capital the investors might venture, which results in more profit returns. On the contrary, under the condition of an accurate prediction, if the degree of belief is little, investors will not put in too much capital, which leads to limited profit. This study attempts to apply belief functions in explaining the prediction results of multivariate fuzzy time series, i.e. the degree of belief that the prediction model has for the prediction result.

9.1 Introduction

In time series analysis, the trend of data can be the basis of detecting events' occurrence such as increasing, decreasing, seasonal cycles or outliers. Hence, by observing certain characteristics, an optimal fitting model can be selected from a prior model family, such as ARIMA models, Threshold models, ARCH models etc. However, there is much information untold in the published (official) data. Take the close price of stock market as an example, how close can the today's record stands for today's trade? How many volumes have been traded at this record? What is the

H.T. Nguyen and B. Wu: *Fundamentals of Statistics with Fuzzy Data*, StudFuzz **198**, 145–182 (2006)
www.springerlink.com

variation that today's trade goes. Under such questions, an attempt to construct a mathematical model via the traditional models and analytical methods to interpret the data and trends of a time series may result in the risk of producing over-fitting models.

One the other hand, the concept of fuzzy sets (logic), first proposed by Zadeh [117], provides a more realistic and moderate approach, by referring to fuzzy measure and classification concept human brain utilizes in dynamic surroundings, to handle the phenomenon of multi-complexities and uncertainties. Because fuzzy theory has intrinsic features of linguistic variables, it can minimize trouble on dealing with uncertain problems. Therefore, fuzzy theory has been widely applied in many fields such as aerospace, mechanical engineering, medical science, power generation, and geology, etc. Among these fields, the application of fuzzy control systems is even more popular; see [57] and [63].

Moreover, it has been widely adopted in social sciences recently. For instance, [20], [111], and [114] presented quantitative methods of approximate reasoning, respectively. [116] proposed fuzzy clustering techniques to construct fuzzy models. [110] applied fuzzy statistics on the analysis of sociology survey. [95], [112].

Recently, the applications of fuzzy logic on dynamic data have caught more and more attention. For example, [9], [68], [85] and [86] used fuzzy theory to construct trend fuzzy time series. [108] proposed a fuzzy identification procedure for ARCH and Bilinear models. [109] used fuzzy statistical techniques in change period's detection of nonlinear time series. [94] proposed a fuzzy ARIMA (FARIMA) model to forecast the exchange rate of NT dollars to US dollars with considering the time-series ARIMA model and fuzzy regression model.

Yet, when applying fuzzy logic in the time series analysis, the first step is considering how to integrate linguistic variables analysis methods to solve problems of data uncertainty. For this, some authors during 1980-1990 proposed a logical examination method with a decision table to describe fuzzy time series models, but this method is hard to be applied to multivariate systems. Thus, to bring more precise fuzzy models, [10], [36] [85], [93] and [115] presented self-learning methods to modify fuzzy models for dynamic system in linguistic field and later [11] proposed a fuzzy linguistic summary as one of the data mining function to discover useful knowledge from database. Besides, using trial-and-error procedure for choosing appropriate weighted factors is still troublesome. In fact, fuzzy relation equations are easier to be understood and applied than decision tables or decision rules.

In view of this, many researchers have adopted fuzzy relation equations for problem solving. For instance, [85] and [86] proposed the procedure for developing fuzzy time series and model theory structure by using fuzzy relation equations. They also later applied this method to forecast enrollments at State University of Alabama. [10] proposed the two-factors time-variant fuzzy time series model and developed two algorithms for temperature prediction, but this method seems lack of statistical modeling and forecasting processing, and moreover, it is not easy to apply their technique into a multivariate systems.

In the humanities and social sciences, fuzzy statistics and fuzzy correlation has gradually got attention. This is a natural result because the complicated phenomenon of humanities and society is hard to be fully explained by traditional models. Regarding stock market as an example, the essence of closing price is uncertain and indistinct. Moreover, there are many factors influence closing price, such as trading volume and exchange rate, etc. Therefore, if we merely consider closing price of yesterday to construct our forecasting model, not only will we misestimate the future trend, but also we will suffer unnecessary loss. While literatures in the past have been focusing on univariate fuzzy time series but lesser on multivariate dynamic data.

In view of this, we propose an integrated procedure for multivariate fuzzy time series modeling and its theory structure through fuzzy relation equations in this research. Furthermore, combining the data of closing price and trading volume, we apply this method to construct multivariate fuzzy time series model for Taiwan Weighted Stock Index and forecast future trend while comparing the forecasting performance by average forecasting accuracy. We strongly believe that this model will be profound of meaning in forecasting future trend of financial market.

9.2 Fuzzy Time Series

The concept of fuzzy logic primarily focuses on people's perception measurement instead of the precise and crisp measurement. In traditional social or economic researches, we often meet the incomplete or uncertain information problem when we are seeking a model family to construct an appropriate time series model for the underlying data. For example, should the number of freshman enrollment be counted from the beginning, middle or end of year? Should we determine the exchange rate of NT dollars to US dollars by opening price, closing price, or the average of price ceiling and price floor?

In this research, we attempt to transform observations into fuzzy sets by using membership functions because they have the features to describe fuzzy sets. This is the very basic concept in fuzzy time series analysis. Through membership functions can we quantify fuzzy sets and further analyze fuzzy information by implementing precise mathematic methods.

The fuzzy time series is a method combining linguistic variables with the analyzing process of applying fuzzy logic into time series to solve the fuzziness of data. Thus, before developing multivariate fuzzy time series model and forecasting, we should give some important definitions for fuzzy time series.

Definition 9.1 Fuzzy time series

Let $\{X_t \in R, t = 1,2,...,n\}$ *be a time series,* Ω *be the range of* $\{X_t \in R, t = 1,2,...,n\}$ *and* $\{ P_i; i = 1,2,..., r, \bigcup_{i=1}^{r} P_i = \Omega \}$ *be an ordered partition on* Ω. *Let* $\{L_i, i = 1,2,...,r\}$ *denote linguistic variables with respect to the ordered partition set. For* $t = 1,2,...,n$, *if* $\mu_i(X_t)$, *the grade of membership of* $\{X_t\}$ *belongs to* L_i, *satisfies* $\mu_i : R \to [0,1]$ *and* $\sum_{i=1}^{r} \mu_i(X_t) = 1$, *then* $\{FX_t\}$ *is said to be a fuzzy time series of* $\{X_t\}$ *and written as*

$$FX_t = \mu_1(X_t)/L_1 + \mu_2(X_t)/L_2 + \cdots + \mu_r(X_t)/L_r,$$

where / is employed to link the linguistic variables with their memberships in FX_t, *and the + indicates, rather than any sort of algebraic addition, that the listed pairs of linguistic variables and memberships collectively.*

For convenience, let us denote FX_t as $FX_t = (\mu_1, \mu_2, ..., \mu_r)$.

When calculating corresponding membership functions of linguistic variables in fuzzy time series, this research uses triangular membership functions for facilitating transformation process.

Example 9.1 Consider the time series $\{X_t\}$ = {0.7, 1.9, 2.7, 4.2, 3.5, 3.1, 4.4, 3.7}. Let $\Omega = [0,5]$ and choose an ordered partition set {[0,1), [1,2), [2,3), [3,4), [4,5]} on Ω. Let $\{L_1, L_2, L_3, L_4, L_5\}$ denote linguistic variables: *Very low=* $L_1 \propto$ [0,1)*; Low=* $L_2 \propto$ [1,2)*; Medium=* $L_3 \propto$ [2,3) *High=* $L_4 \propto$ [3,4)*; Very high=* $L_5 \propto [4,5]$.

We evaluate the mean $\{m_1 = 0.5, m_2 = 1.5, m_3 = 2.5, m_4 = 3.5, m_5 = 4.5\}$ of the ordered partition set. Since X_1 is between 0.5 and 1.5, and

$$\frac{1.5-0.7}{1.5-0.5} = 0.8 \in L_1 \ , \ \frac{0.7-0.5}{1.5-0.5} = 0.2 \in L_2 \, ,$$

we get the fuzzy set FX_1 with respect to X_1 is (0.8,0.2,0,0,0). Similarly, we can get the following Table 91.

Table 9.1 Fuzzy time series $\{FX_t\}$ of $\{X_t\}$

	Very low	Low	Medium	High	Very high
$FX_1 =$	0.8	0.2	0	0	0
$FX_2 =$	0	0.6	0.4	0	0
$FX_3 =$	0	0	0.8	0.2	0
$FX_4 =$	0	0	0	0.3	0.7
$FX_5 =$	0	0	0	1	0
$FX_6 =$	0	0	0.4	0.6	0
$FX_7 =$	0	0	0	0.1	0.9
$FX_8 =$	0	0	0	0.8	0.2

In time series modeling and analysis, the determination of autocorrelation value is very important because it reflects the autocorrelation trend of a time series. But for a set of uncertain or incomplete data, its autocorrelation should not be explained with a number. Therefore, this research attempts to use fuzzy relation to analyze the autocorrelation level in the fuzzy time series. Fuzzy relation is also a key point to be explored in the research of fuzzy time series.

Definition 9.2 Relation of fuzzy sets

Let $\{P_i, i = 1,2,\ldots,r\}$ be an ordered partition set on Ω, $G = (\mu_1,\ldots,\mu_r)$ and $H = (\nu_1,\ldots,\nu_r)$ be fuzzy sets, then a fuzzy relation R between

G and H is $R = G^t \circ H = [R_{ij}]_{r \times r}$ *,where* μ_i, ν_j *denote memberships,* t *denotes transpose, and* $R_{ij} = \min(\mu_i, \nu_j)$.

Fuzzy Markov Relation Matrix *R*

We can find that fuzzy relation is the key for constructing good fuzzy time series models. If we can precisely handle fuzzy relation matrix through fuzzy relation, then fuzzy time series models will provide a better fitting result. Besides, there are many different ways for calculating a fuzzy relation matrix. Dubois and Prade (1991), Wu (1986) had proposed some methods to calculate fuzzy relation matrix but none of them is based on the same premises.

In this research, we consider multivariate fuzzy time series under the assumption that the time series trend for each variable is stationary. This is because most of the operations in financial market conform to the Markov property. Therefore we will focus on exploring multivariate fuzzy time series with Markov property. For this, we would begin by defining fuzzy Markov relation matrix $\Re$ before constructing multivariate fuzzy time series models,

Definition 9.3 Fuzzy Markov relation matrix

Assume that $\{ FX_t, t = 1,2,...,n \}$ *is a FAR (1) (fuzzy autoregression of order one) fuzzy time series, i.e. for any time* t *,* FX_t *depends only on* FX_{t-1} *. If the fuzzy set* FX_t *consists of finite membership functions* $\mu_i(X_t), i = 1,2,....,r$ *, then*

$$\boldsymbol{R} = [R_{ij}]_{r \times r} = \max_{2 \le t \le n}[\min(\mu_i(X_{t-1}), \mu_i(X_{t-1}))]_{r \times r}$$

is called a fuzzy-Markov-relative matrix.

Example 9.3 Consider a fuzzy time series $\{FX_t\}$ from Table 9.1 of Example 9.1. Suppose that this fuzzy time series $\{FX_t\}$ is an autoregressive process of order one and detect the linguistic variable according to the position of the greatest membership. We can find the relationships with linguistic variables for this fuzzy time series as follow: $L_1 \to L_2$; $L_2 \to L_3$; $L_3 \to L_5$; $L_5 \to L_4$; $L_4 \to L_4$; $L_4 \to L_5$; $L_5 \to L_4$.

Since L_1 and L_2 represent (1,0.5,0,0,0) and (0.5,1,0.5,0,0), respectively. By Definition 2.2, we get the fuzzy relation of $L_1 \to L_2$ is

$$R_1 = \begin{bmatrix} 1 \\ 0.5 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0.5 & 1 & 0.5 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.5 & 1 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Similarly, we can get the fuzzy relations of $\{FX_t\}$by:

$$R_1 = \begin{bmatrix} 0.5 & 1 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, R_2 = \begin{bmatrix} 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$R_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad R_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 1 & 0.5 \end{bmatrix},$$

$$R_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0.5 \end{bmatrix}, \quad R_6 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix},$$

$$R_7 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 1 & 0.5 \end{bmatrix}.$$

Finally, by Definition 2.3, we get the following fuzzy Markov relation matrix.

$$\boldsymbol{R} = \begin{bmatrix} 0.5 & 1 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 & 1 \\ 0 & 0 & 0.5 & 1 & 1 \\ 0 & 0 & 0.5 & 1 & 0.5 \end{bmatrix}.$$

9.3 Multivariate Fuzzy Time Series Modeling and Forecasting

In multivariate fuzzy time series analysis, the data can be numerical or qualitative formats, or linguistic values (such as data derived from tasting). How to participate the fuzzy data into several ranks is a prior problem. There is no definite rule about the suitable number of order partition in building fuzzy range sets. In general, the more the number of order partition, the higher the precision will be. But it will take more time in computing processing and more complicated form in modeling.

Though there is no certain method to determine canonical value in the process of fuzzification, we think that the typical value can be derived from the median, mean or cluster center of all elements in each "*point*". However, a drawback of using cluster center as typical value is that several typical values may appear in a given set, making transformation process more complicated if the set is a trapezoidal membership function. On the other hand, using median or the mean as typical value would only create one typical value in every given set, making the transformation process more straightforward if the set is a triangular membership function.

In the issues we assume the fuzzy time series is kind of stationary. However, if the collected data are fuzzy data transformed from numerical data with increasing or decreasing trend, difference should be used to stabilize original data and then modeling can be continued.

Order identification plays an important role in the multivariate fuzzy time series analysis. If we can get an accurate order, the factors influencing trend can be grasped and practical mathematical models can be built. In practical application, for stocks index, exchange rates etc, are usually with nonlinear characteristics, high prices tend to be higher and low prices tend to be lower. Also, most of the operations in financial markets conform to the Markov property. Therefore, the choice of order can depend on real situation.

Once the order of multivariate fuzzy time series model is determined, the method mentioned in Section 9.2, together with multiple factors, can be used to obtain a fuzzy Markov relation matrix. Though Lee (1994) has used real fuzzy outputs and estimated fuzzy outputs to get smaller error, this method involves too many numerical operations and is therefore too complicated. Hence, this research adopts the method proposed by Wu (1986), Sugeno and Tanaka (1991) to calculate fuzzy Markov relation matrix $\boldsymbol{R}$.

Having above explanations, we start to analyze the multivariate fuzzy time series model. Here we only focus on the research of multivariate fuzzy autoregressive time series of order one model.

Definition 9.4 The FVAR(1) time series

If a multivariate fuzzy time series, $\{(FX_{1t}, FX_{2t}, ..., FX_{kt})\}$ *for any t, can be written as*

$$(FX_{1t}, FX_{2t}, ..., FX_{kt}) = (FX_{1t-1}, FX_{2t-1}, ..., FX_{kt-1})\begin{bmatrix} R_{11} & \cdots & \cdots & R_{1k} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ R_{k1} & \cdots & \cdots & R_{kk} \end{bmatrix}$$

where R_{ij} *is the fuzzy-Markov-relative matrix for* i^{th} *variable relative to* j^{th} *variable,* $i, j = 1,2,...,k$ *, then we call the multivariate fuzzy time series* $\{(FX_{1t}, FX_{2t}, ..., FX_{kt})\}$ *a multivariate first order of auto-regressive fuzzy time series model and denote it as VFAR(1). In the model,* $(FX_{1t}, FX_{2t}, ..., FX_{kt})$*, is only depends on* $(FX_{1,t-1}, FX_{2,t-1}, \cdots, FX_{k,t-1})$ *, thus the model can be referred as the Markov Process.*

Principle of qualitative identification by fuzzy rule base

In the multivariate fuzzy time series, one of the most important points is how to transform fuzzy numbers (membership functions) into corresponding linguistic variables (attributions). In general, corresponding linguistic variables is determined by the greatest membership function. However, if there is more than one greatest membership function, how to make a decision among them? On the other hand, there are many factors influencing the underlying time series, for instance the Taiwan Weighted Stock Index, such as the trading volume, exchange rates, interest rates, political causes, and so on.

However, in a real-time and open markets, the closing price of previous day maybe the first important factor to determine the next period of price. The second important factor maybe the trading volume. Because trading volume usually stands for the leading indicator of stock price, the increasing trading volume is a prerequisite for increasing stock price. Besides, in the traditional time series analysis, there is high autocorrelation between previous data and current data. These two characteristics mentioned above can thus be used for constructing multivariate fuzzy time series. Therefore, through fuzzy Markov relation matrix, we can obtain current fuzzy numbers from previous fuzzy numbers. The question here is how to transfer fuzzy numbers to corresponding linguistic variables? First, let's give the following definition,

Definition 9.5 Linguistic vector indicator function

Let $L_i = \{(L_{i1},\cdots,L_{ir}); L_{ij}$ *: linguistic variable;* $j = 1,\ldots,r\}$ *be a linguistic vector of* $\{FX_{i,t}\}$ *and* $F\hat{X}_{i,t}$ *be a vector of memberships in* L_i, $i = 1,\ldots,k$. *Then* $F\widetilde{X}_{i,t} = \{(I_{i1_t},\cdots,I_{ir_t})\,;\, I_{ij_t} = 1$ *or 0;* $j = 1,\ldots,r\}$ *is said to be linguistic vector indictor function,* $i = 1,\ldots,k$, *and*

$$I_{ij_t} = \begin{cases} 1, & if\ \ \mu_{L_{ij}}(F\hat{X}_{i,t}) \ge k \\ 0, & if\ \ \mu_{L_{ij}}(F\hat{X}_{i,t}) < k \end{cases},$$

where $\mu_{L_{ij}}(F\hat{X}_{i,t})$ *denote membership of* $F\hat{X}_{i,t}$ *in* L_{ij}.

Example 9.4

Let $[L_1,L_2] = \{[(L_{11},L_{12},L_{13},L_{14},L_{15}),(L_{21},L_{22},L_{23},L_{24},L_{25})]\}$; where $L_{11} = plunge$, $L_{12} = drop$, $L_{13} = draw$, $L_{14} = soar$, $L_{15} = surge$; $L_{21} =$ *very low*, $L_{22} = low$, $L_{23} = medium$, $L_{24} = high$, $L_{25} =$ *very high*$\}$ be a bivariate linguistic vector of $\{(FX_{1,t},FX_{2,t})\}$. After calculating a bivariate time series data by fuzzy Markov relation matrix, we have $[F\hat{X}_{1,t}, F\hat{X}_{2,t}] = [(1,1.5,2,2,1.5),(1,1.5,1.5,2,1.5)]$. By Definition 3.2, we get $[F\widetilde{X}_{1,t},F\widetilde{X}_{2,t}]$=[(0,0,1,1,0),(0,0,0,1,0)].

According to Definition 9.2, it is easy to see that we can transfer fuzzy numbers from the multivariate fuzzy time series models to linguistic vector indicator functions. In this section, we use Definition 9.2 and establish Threshold function by fuzzy reasoning to obtain a fuzzy rule base and further analyze its outputting linguistic variables.

The fuzzy rule base is exactly expert systems (rules) established to deal with some fuzzy phenomenon or knowledge in daily life. The rule in this research is based on the fact that traditional times series itself is statistical dependent and then we use autocorrelation function (ACF) and partial autocorrelation function (PACF) to find the coefficients of time series models. Since the ACF and PACF are not clear in nonlinear time series, we can follow traditional AutoRegressive Integrated Moving Average (ARIMA) models and use three constructing steps: (1) Order identification, (2) Parameter estimation and (3) Diagnostic checking to help us get optimal fitting models. Hence, the fuzzy rule established in this research is basically obtained from above concept, accumulated experience, and fuzzy reasoning. So this method of construction is intuitive and subjective.

According to the times series applied in this research, we build ranges as {plunge, drop, draw, soar, surge} and {very low, low, medium, high, very high} for closing price and trading volume difference and thus clustering the value of n into 5 parts. We also use $(I_{k1_t}, \cdots, I_{k5_t})$ as fuzzy inference indicator, where $I_{kj_t} = 1$ or 0 and $j = 1,2,...,5$, and thus 32 linguistic vector indicator functions can be established. Yet, we need to exclude vector (0,0,0,0,0) because it cannot represent any linguistic variable. However, it’s not easy to categorize 31 linguistic vector indicator functions to their corresponding linguistic variables. If only a “1” appears in the linguistic vector indicator function, the output will be the corresponding linguistic variable where this 1 is located. For example, (0,0,0,1,0) represents that the membership function of "soar" is "1", so the outputting of linguistic variable is "soar."

How to deal with the scenario when there is more than one "1" appearing in the linguistic vector indicator function? It will be very time-consuming if we attempt to sequentially examine each component of linguistic vector indicator function in fuzzy time series, i.e. examining each I_{kj_t}, where $j = 1,2,...,5$, from I_{k1_t} = 1 or 0 until I_{k5_t} = 1 or 0. Yet, if we take entire linguistic vector indicator function to judge, it will be easy for us to identify its representative linguistic variable through experience rules. For example, (0,0,0,1,1) represents both the membership functions of "soar" and "surge" are 1. Through experience rules, we can detect its outputting linguistic variable is "surge." Similarly, (1,1,0,0,0) represents that the outputting linguistic variable is "plunge."

Using the above methodology, we consider the threshold function H_t as our decision process for different range partition sets, where H_t is defined as follows:

$$H_t = \begin{cases} plunge\ (very\ low), & if\ K_t \le -2 \\ drop\ (low), & if\ -2 < K_t \le -1,\ or\ if\ K_t = -2\ and\ \sum_{j=1}^{5} I_{kj_t} \ge 3 \\ draw\ (medium), & if\ K_t = 0 \\ soar\ (high), & if\ 1 \le K_t < 2,\ or\ if\ K_t = 2\ and\ \sum_{j=1}^{5} I_{kj_t} \ge 3 \\ surge\ (very\ high), & if\ K_t \ge 2 \end{cases}$$

where $K_t = \sum_{j=1}^{5} (j-3) I_{kj_t}$.

Finally, we can use this threshold function H_t to establish the following fuzzy rule base.

Fuzzy rule base

For $i = 1,...,k$,

(1) If $\widetilde{FX}_{i,t} \in$ *{(1,0,0,0,0), (1,1,0,0,0), (1,0,1,0,0), (1,1,1,0,0)}, then the outputting linguistic variable is" plunge (very low) "*

(2) If $\widetilde{FX}_{i,t} \in$ *{(0,1,0,0,0), (1,1,0,1,0), (1,1,1,0,1), (1,1,0,0,1), (1,0,0,1,0), (1,1,1,1,0), (0,1,1,0,0), (1,0,1,1,0)}, then the outputting linguistic variable is" drop (low) "*

(3) If $\widetilde{FX}_{i,t} \in$ *{(0,0,1,0,0), (1,0,1,0,1), (1,0,0,0,1), (1,1,1,1,1), (0,1,0,1,0), (1,1,0,1,1), (0,1,1,1,0)}, then the outputting linguistic variable is" draw (medium) "*

(4) If $\widetilde{FX}_{i,t} \in$ *{(0,0,0,1,0), (0,1,0,1,1), (1,0,1,1,1), (1,0,0,1,1), (0,1,0,0,1), (0,1,1,1,1), (0,0,1,1,0), (0,1,1,0,1)}, then the outputting linguistic variable is" soar (high) "*

(5) If $\widetilde{FX}_{i,t} \in$ *{(0,0,0,0,1), (0,0,0,1,1), (0,0,1,0,1), (0,0,1,1,1)}, then the outputting linguistic variable is " surge (very high)"*

In Example 3.2, we obtained $[\widetilde{FX}_{1,t}, \widetilde{FX}_{2,t}]$ =[(0,0,1,1,0),(0,0,0,1,0)]. Therefore, through fuzzy rule base mentioned above, we can get the outputting linguistic variables for price limit and trading volume difference are "soar" and "high", respectively.

Multivariate fuzzy time series forecasting

For the multivariate fuzzy autoregressive process of order one model

$$(FX_{1,t}, FX_{2,t}, \ldots, FX_{k,t}) = (FX_{1,t-1}, FX_{2,t-1}, \ldots, FX_{k,t-1}) \begin{pmatrix} R_{11} & \cdots & \cdots & R_{1k} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ R_{k1} & \cdots & \cdots & R_{kk} \end{pmatrix}.$$

and observations $(FX_{1,t}, FX_{2,t}, \ldots, FX_{k,t})$, $t = 1,2,\ldots,n$, *then*

(1) One-step prediction is

$$(FX_{1,n}(1), FX_{2,n}(1), \ldots, FX_{k,n}(1)) = (FX_{1,n}, FX_{2,n}, \ldots, FX_{k,n}) \begin{pmatrix} R_{11} & \cdots & \cdots & R_{1k} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ R_{k1} & \cdots & \cdots & R_{kk} \end{pmatrix}.$$

(2) Two-step prediction is

$$(FX_{1,n}(2), FX_{2,n}(2), \ldots, FX_{k,n}(2)) = (FX_{1,n}, FX_{2,n}, \ldots, FX_{k,n}) \begin{pmatrix} R_{11} & \cdots & \cdots & R_{1k} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ R_{k1} & \cdots & \cdots & R_{kk} \end{pmatrix}^2$$

(3) l *-step prediction is*

$$(FX_{1,n}(l), FX_{2,n}(l), \ldots, FX_{k,n}(l)) = (FX_{1,n}, FX_{2,n}, \ldots, FX_{k,n}) \begin{pmatrix} R_{11} & \cdots & \cdots & R_{1k} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ R_{k1} & \cdots & \cdots & R_{kk} \end{pmatrix}^l.$$

Average accuracy of forecasting

Once multivariate time fuzzy series model and fuzzy rule base are established, to compare the error between outputting linguistic variables and real ones, this research use the concept of non-parametric rank to provide corresponding value to each linguistic variable. For instance, a "plunge" as -2, "drop" as -1, "draw" as 0, "soar" as 1 and "surge" as 2. By doing so, the average forecasting accuracy can be defined.

Definition 9.7 Average accuracy of forecasting

Suppose $\{RL_t, t=1,\cdots,n\}$ and $\{FL_t, t=1,\cdots,n\}$ denote the real and outputting linguistic variables of the time series, respectively. Let $L=$ { $(L_1, L_2, \cdots, L_r)$ = $(-(r-1)/2, -(r-3)/2, \cdots, (r-1)/2)$; L_j : linguistic variable} be corresponding values of linguistic variables, then P is said to be average accuracy of forecasting and written as

$$P = 1 - \frac{1}{n} \sum_{t=1}^{n} \frac{|FL_t - RL_t|}{r-1}.$$

where r denote the number of linguistic variables.

Example 9.5 Suppose that real linguistic variables of the time series are {drop, draw, drop, surge, draw, drop, surge, drop, draw, plunge}, then the corresponding values of linguistic variables are {-1 , 0 , -1 , 2 , 0 , -1 , 2 , -1 , 0 , -2}. The outputting linguistic variables are {drop, draw, plunge, surge, draw, draw, surge, drop, surge, draw}, then the corresponding values of linguistic variables are {-1 , 0 , -2 , 2 , 0 , 0 , 2 , -1 , 2 , 0}. By Definition 3.4, we can get

$$P = 1 - \frac{1}{10} \sum_{t=1}^{10} \frac{|FL_t - RL_t|}{4} = 1 - \frac{6}{40} = 0.85.$$

We further provide the integrated process and flowchart for multivariate fuzzy time series modeling,

An integrated process for multivariate fuzzy time series modeling

Step 1: Observe time series $\{X_{1t}\},\cdots,\{X_{kt}\}$*. Decide the range* Ω_i *and the linguistic variables* $\{L_{i1},L_{i2},\cdots,L_{ir}\}$ *of* $\{X_{it}\}$, $i=1,2,\cdots,k$.

Step 2: Calculate the fuzzy time series $\{FX_{it}\}$ *of* $\{X_{it}\}$, $i=1,2,\cdots,k$ *and detect the linguistic variable according to the position of the greatest membership in* $FX_{it}, t=1,2,...,n$.

Step 3: Calculate the fuzzy relations between $\{FX_{it}\}$ *and* $\{FX_{jt}\}$, $i,j=1,2,\cdots,k$.

Step 4: By Step 3, according to all fuzzy relations between $\{FX_{it}\}$ *and* $\{FX_{jt}\}$, $i,j=1,2,\cdots,k$*, we can get the fuzzy Markov relation matrix, then constructing a multivariate fuzzy time series model.*

Step 5: Examining $F\widetilde{X}_{i,t}$, $i=1,2,\cdots,k$*. If the number of "1" is only one, we can detect the corresponding linguistic variable immediately, otherwise detect the corresponding linguistic variables by fuzzy rule base.*

Step 6：Forecasting according to the multivariate fuzzy time series model.

Step 7：Stop.

9.4 Measuring Beliefs in the Forecasting Process

As people attempt to make decision for future affairs, they usually refer to past experience. Multivariate fuzzy time series forecasting model is just like human decision-making model. Fuzzy relative matrix derives from previous quantity-price relationship and is just like former experience imprinted in human brain. Thus, by means of fuzzy relative matrix, we can further achieve prediction results. Nonetheless, just like people's judgments, people are not always right about the decisions they make. The same goes to the results of prediction made by forecasting models. Thereupon comes the question: How should we regard the prediction results forecasted by the model each time? Should we totally, partially or hardly accept them?

Try to recall the reaction we had when we had to make decisions for the future. Obviously, there is more than just the decision itself. The degree of belief towards the decision is also there. In other words, people not only make decisions but also make decisions with "degree of belief" in them. They tend to execute decisions more bravely for confident ones. By contrast, they act more carefully for less confident decisions. Accordingly, in the case of trying to make an investment, if we have faith in future market, we can make a large sum of investment; whereas, if we have little faith in future market, we can reduce the sum of money to be invested. In this regard, we can avoid loss due to misjudgment, increase investment returns for right judgment, and enhance our risk-control ability for capital.

So, this study not only makes predictions on multivariate fuzzy time series, but also establishes belief functions. Belief functions are used to describe the degree of belief in the predictions we made. Since the theoretically mathematical foundation is constructed on the basis of the notion of belief function, some definitions and properties of belief function are provided, under limited conditions, as following:

Definition 9.8 *Let* Ω *be a finite set. The set function* $G : 2^{\Omega} \to [0,1]$ *is a belief function, if*

(i) $G(\phi) = 0, G(\Omega) = 1$.

(ii) For any $n \geq 1$ *and* $A_i \subseteq \Omega$, $i = 1,\ldots,n$,

$$G\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{\phi \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} G\left(\bigcap_{j \in I} A_j\right).$$

Theorem 9.1

(i) Let G be a belief function in Ω. *Then the function* g *defined on* 2^{Ω}, $g(A)=\sum_{B\subseteq A}(-1)^{|A-B|}G(B)$, *is nonnegative.*

(ii) Let G and g be two functions from 2^{Ω} *to* R. $G(A)=\sum_{B\subseteq A}g(B)$ *if and only if* $g(A)=\sum_{B\subseteq A}(-1)^{|A-B|}G(B)$.

Proof.

(i) Let $A=\{w_1,w_2,...,w_n\}\subseteq\Omega$ and $A_i=A-\{w_i\}$. Then

$$G(A)\geq\sum_{i=1}^{n}G(A_i)-\sum_{i<j}G(A_i\cap A_j)+...+(-1)^n\sum_{i=1}^{n}G\left(\bigcap_{j\neq i}A_j\right),$$

Note: $\bigcap_{i=1}^{n}A_i=\phi$, so $g(A)\geq 0$.

(ii) "$\Rightarrow$"Assume $G(A)=\sum_{B\subseteq A}g(B)$. Then

$$\sum_{B\subseteq A}(-1)^{|A-B|}G(B)=\sum_{B\subseteq A}(-1)^{|A-B|}\sum_{D\subseteq B}g(D)=\sum_{D\subseteq B\subseteq A}(-1)^{|A-B|}g(D).$$

If $D = A$, then the last form is $g(A)$. If $D\neq A$, $A-D$ has $2^{|A-D|}$ subsets. Therefore, there exist even number of subsets B, $D\subseteq B\subseteq A$, exactly half of the sets have even number of elements. Thus the number of +1 is $(-1)^{|A-B|}$, half of the total number and the other is -1. For each D, $D\neq A$,

$$\sum_{D\subseteq B\subseteq A}(-1)^{|A-B|}g(D)=0. \text{ So } \sum_{B\subseteq A}(-1)^{|A-B|}G(B)=g(A).$$

"$\Leftarrow$"The way is the same as above.

To explain Theorem 3.1, we discover that any belief function G in Ω can be written as $g:2^{\Omega}\to[0,1]$, $\sum_{A\subseteq\Omega}g(A)=1$, and $g(\phi)$ =0. Hence, formally g is a probability density function of some random set S in Ω, i.e.

P(S=A)=g(A), and $G(A)=P(S\subseteq A)$. Refer to the Theorem 3.2 below. Besides G plays the role of the distribution function of the random set S. In the theory of evidence, the value $g(A)$ is interpreted as the weight of evidence in support of A.

Theorem 9.2 *Let* $g:2^{\Omega}\to[0,1]$, $g(\phi)=0$ *and* $\sum_{A\subseteq\Omega}g(A)=1$. *Define* $G(A)=\sum_{B\subseteq A}g(B)$, *then the set function G is a belief function.*

Proof. *Since* $G(\phi)=0$ *and* $G(\Omega)=1$, *we have to show that G is infinitely monotone. Let I = {1, 2, .., n} and* $A_i\subseteq\Omega$, $i\in I$. *We have*

$$G\left(\bigcup_I A_i\right)=\sum_{B\subseteq\bigcup_I A_i}g(B)\geq\sum_{B\in\Gamma}g(B),$$

where at least one A_i *contains* Γ *which is a subcollection of* $\bigcup_I A_i$.

$$\sum_{\phi\neq J\subseteq I}(-1)^{|J|+1}G\left(\bigcap_J A_j\right)=\sum_{\phi\neq J\subseteq I}(-1)^{|J|+1}\sum_{B\subseteq\bigcap_J A_j}g(B)$$

$$=\sum_{B\in\Gamma}\sum_{\phi\neq J\subseteq I}(-1)^{|J|+1}g(B)=\sum_{B\in\Gamma}g(B).$$

Since $\sum_{\phi\neq J\subseteq I}(-1)^{|J|+1}$ *=1, so the theorem is proved.*

How to establish and calculate the degree of belief ?

This study uses the maximum membership grade to convert fuzzy forecasting value into linguistic vector index, and then obtain the predicted attributes of linguistic variables upon the fuzzy rule base. Just as when people are not sure to the very same extent if they make the right decision for the future, multivariate time series model doesn't have the same degree of belief in every prediction it makes. Thus, it is necessary to form a function to evaluate the degree of belief of the forecasting model. In order to form and calculate degree of belief, the following definitions must be given:

Definition 9.9 Generalized membership rank and maximum membership

Let $L=\{(L_{11},\cdots L_{15}),\cdots,(L_{k1},\cdots L_{k5}):L_{ij}\ \textit{is a linguistic}\ \text{var}\textit{iable}\}$, *and* FX_t *be the membership function of multivariate fuzzy time series with respect to L. Let* $F\overline{X}_t$ *be the linguistic vector index converted by* FX_t. *That is* $F\overline{X}_t=\{(I_{11_t},\cdots I_{15_t}),\cdots,(I_{k1_t},\cdots I_{k5_t}),I_{ij_t}=1\ or\ 0\}$.

Assume that FA_{ti} *is the generalized membership rank of* FA_{ti} *and* FC_{ti} *is the maximum membership of* $F\hat{X}_{ti}$, *where*

$$FA_{ti}=\left[\sum_{j=1}^{5}(j-3)I_{ij_t}\right]\Big/\left[\sum_{j=1}^{5}I_{ij_t}\right],$$

$$FC_{ti}=\max\left[\mu_{L_{i1}}(FX_t),\cdots,\mu_{L_{i5}}(FX)\right],$$

in which $\mu_{L_{ij}}(FX_t)$ *is the membership function of* FX_{ti} *in the linguistic variable* L_{ij}.

Example 9.6 *Assume that there is a membership function of multivariate fuzzy time series with respect to L, say*

FX_t = {0.56, 0.95, 1.15, 1.38, 1.38}, (0.96, 1.32, 1.40, 0.74, 0.28) }. *Then* $F\overline{X}_t=\{(0,0,0,1,1),(0,0,1,0,0)\}$. *According to definition 9.2, we have*

$FA_{t1}=(-2\times0-1\times0+0\times0+1\times1+2\times1)/(0+0+0+1+1)=1.5$

$FA_{t2}=(-2\times0-1\times0+0\times1+1\times0+2\times0)/(0+0+1+0+0)=0$.

We computed the maximum membership, $FC_{t1}=1.38$ ， $FC_{t2}=1.40$.

Definition 3.3 Confidence interference grade

For each $F\hat{X}_t$, *let* FI_{tij} *be the confidence interference grade of each element for generalized membership rank in* $F\hat{X}_{ti}$.
Then

$$FI_{tij} = \frac{(1-I_{ij})\mu_{L_{ij}}(FX_t)}{FC_{ti}},\ j=1,2,\ldots 5$$

where $\mu_{L_{ij}}(FX_t)$ *is the membership function of* FX_t *in the linguistic variable* L_{ij}.

Example 9.7

Assume that $F\hat{X}_t = \{(0.56,0.45,1.15,1.38,1.38),(0.96,1.32,1.40,0.74,0.28)\}$.
Then the Confidence interference grade of each element is as following:

$FI_{t11} = 0.41$, $FI_{t12} = 0.33$, $FI_{t13} = 0.83$, $FI_{t14} = 0$, $FI_{t15} = 0$.
$FI_{t21} = 0.66$, $FI_{t22} = 0.94$, $FI_{t23} = 0$, $FI_{t24} = 0.53$, $FI_{t25} = 0.2$.

Definition 9.10 The weight of confidence interference grade

For the confidence interference grade of each element in $F\hat{X}_{ti}$, *let* FW_{tij} *be the weight of confidence interference grade of* FI_{tij}, *then*

$$FW_{tij} = \frac{|(j-3)-FA_{ti}|}{4},\ j=1,2,\ldots,5$$

Example 9.8

Assume $FX_t = \{(0.56,0.45,1.15,1.38,1.38)(0.96,1.32,1.40,0.74,0.28)\}$, *then the weight of confidence interference grade of each element is as following respectively*

$FW_{t11} = 0.88$, $FW_{t12} = 0.63$, $FW_{t13} = 0.38$, $FW_{t14} = 0.13$, $FW_{t15} = 0.13$.
$FW_{t21} = 0.5$, $FW_{t22} = 0.25$, $FW_{t23} = 0$, $FW_{t24} = 0.25$, $FW_{t25} = 0.5$.

Definition 9.11 Belief function

Let $L = \{(L_{11}, \cdots L_{15}), ..., (L_{k1}, \cdots L_{k5}), L_{ij}$ *is a linguistic* var*iable*$\}$ *and* FX_t *be a membership function of multivariate fuzzy time series with respect to* L. *Let* C_{ti} *be the belief function of* FX_{ti} , *then*

$$C_{ti} = 1 - \frac{\sum_{j=1}^{5} FI_{tij} \cdot FW_{tij}}{\sum_{j=1}^{5} (1 - I_{ij})},$$

where $FX_t = \{FX_{t1}, FX_{t2}, ..., FX_{tk}\}$

Example 9.9

Assume that
$FX_t = \{(0.56, 0.45, 1.15, 1.38, 1.38)(0.96, 1.32, 1.40, 0.74, 0.28)\}$. *Then the belief functions are*

$$C_{t1} = 1 - \frac{0.41 \times 0.875 + 0.33 \times 0.625 + 0.83 \times 0.375 + 0.125 \times 0 + 0.125 \times 0}{(1 + 1 + 1 + 0 + 0)} = 0.71$$

and

$$C_{t2} = 1 - \frac{0.66 \times 0.5 + 0.94 \times 0.25 + 0 \times 0 + 0.53 \times 0.25 + 0.2 \times 0.5}{(1 + 1 + 0 + 1 + 1)} = 0.80.$$

Property 9.1 *Let* C_{ti} *be the belief function of fuzzy forecasting value* FX_{ti}. *For generalized rank* FA_{ti} *and maximum membership* FC_{ti}, *if the confidence interference grade and the weight coefficient in each element get smaller, then the confidence function* C_{ti} *gets higher.*

Property 9.2 *If the distribution pattern of membership in* FX_{ti} *is of single kurtosis, then the larger the maximum membership value* FC_{ti} *is, the higher the belief function value will be.*

Property 9.3 *If the distribution pattern of membership is approximately uniform distribution, then the belief function value* C_{ti} *will be lower.*

Property 9.4 *If the forecasting property is unchanged (medium), then the belief function value* C_{ti} *will be higher. However, if the forecasting property is soar (very high) or plunge (very low), then the belief function value* C_{ti} *will be lower.*

Property 9.5 *Assume that* $C_{(t-1)i}$ *and* C_{ti} *are the belief functions of fuzzy forecasting values* $FX_{(t-1)i}$ *and* FX_{ti} *respectively. Then* $C_{(t-1)i}$ *will not affect* C_{ti}. *In other words, the forecasting value of belief function of a certain day does not affect the one in the next day.*

9.5 Empirical Studies

Data Analysis

The data in this study are the daily weighted stock price index fluctuation and trade volume high/low difference information, taken from Taiwan Stock Exchange Corporation from 2003 January 3 to 2003 March 11, as illustrated in Figures 9.1 and 9.2.

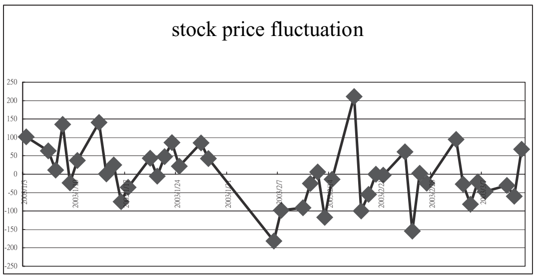


Figure 9.1 TAIEX Fluctuation Summary（2003/January/3 ~ 2003/March/11）

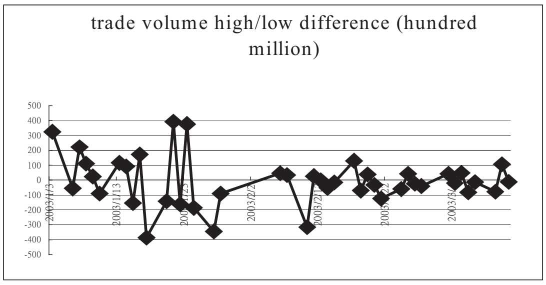


Figure 9.2 TAIEX Trade Volume High/Low（2003/January/3 ~ 2003/March/11）

As shown in the data, the daily maximum value of stock price index fluctuation is 211.09 and the daily minimum one, –181.58; the maximum value of daily trade volume high/low difference is 393(hundred million), and the minimum one, -387(hundred million). Generally speaking, universe of discourse should include the maximum and the minimum values; hence, we choose set (-181.58, 211.09) and (-387, 393) as the universe of discourse of everyday stock price index fluctuation and daily trade volume high/low difference respectively. Because this study is based on fuzzy theory, we need to first fuzzify the data and, then, proceed to establish the model. Thus, (-181.58, 211.09) and (-387, 393) are divided 5 intervals as follows: let E_1=1/8 quantile, E_2=3/8 quantile, E_3=5/8 quantile, E_4=7/8 quantile.

I_{11} =*(minimum, E_1)=(-181.58,-91.5), and its representative value is -181.58*；

I_{12} =*(E_1, E_2)=(-91.5,-24.23), and its representative value is -46.58*；

I_{13} =*(E_2, E_3)=(-24.23,11.22), and its representative value is -2.35*；

I_{14} =*(E_3, E_4)= (11.22,85.53), and its representative value is 43.25*；

I_{15} =*(E_4, maximum)=(85.53,211.09), and its representative value is 211.09.*

I_{21} =*(minimum, E_1)=(-387,-154), and its representative value is -387*；

I_{22} =*(E_1, E_2)=(-154,-52), and its representative value is -81*；

I_{23} =*(E_2, E_3)=(-52, 34), and its representative value is -16*；

I_{24} =*(E_3, E_4)=(34,129), and its representative value is 48*；

I_{25} =*(E_4, maximum)=(129,393), and its representative value is 393*

Among which, $\{I_{11}, I_{12}, I_{13}, I_{14}, I_{15}\}$ and $\{I_{21}, I_{22}, I_{23}, I_{24}, I_{25}\}$ are the five intervals of (-181.58, 211.09) and (-387, 393) respectively. Then, we define five linguistic variables within (-181.58, 211.09) and (-387, 393), that is, $L_{11} = plunge$ ； $L_{12} = down$ ； $L_{13} = unchanged$ ； $L_{14} = up$ ； $L_{15} = soar$.

$L_{21} =$ very low； $L_{22} =$ low； $L_{23} =$ medium； $L_{24} =$ high； $L_{25} =$ very high. Each of the linguistic variables stands for a fuzzy set, and the components of each fuzzy set are I_{ij} ($i = 1,2; j = 1,2,...,5$) and the corresponding membership function.

Establishment of Fuzzy Time Series Analysis Model

Before constructing the model, we have to fuzzify both daily weighted stock price index fluctuation and trade volume high/low difference. By applying the procedure of Definition 2.1, for each fuzzy set $L_{ij}\left(i=1,2;\ j=1,2,\cdots 5\right)$, we gain daily weighted stock price index fluctuation and trade volume high/low difference as well as the corresponding membership function of each linguistic variable, as shown in Table 9.1 and 9.2. For the brief reason we only illustrated the first ten data

Table 9.1 Membership Grade of Daily TAIEX Fluctuation

Date	*Daily TAIEX Fluctuation*	L_{11}	L_{12}	L_{13}	L_{14}	L_{15}
2003/1/3	101.45	0.00	0.00	0.00	0.65	0.35
2003/1/6	63.54	0.00	0.00	0.00	0.88	0.12
2003/1/7	11.22	0.00	0.00	0.70	0.30	0.00
2003/1/8	135.85	0.00	0.00	0.00	0.45	0.55
2003/1/9	-23.2	0.00	0.47	0.53	0.00	0.00
2003/1/10	37.07	0.00	0.00	0.14	0.86	0.00
2003/1/13	140.46	0.00	0.00	0.00	0.42	0.58
2003/1/14	1.16	0.00	0.00	0.92	0.08	0.00
2003/1/15	25.28	0.00	0.00	0.39	0.61	0.00
2003/1/16	-74.41	0.20	0.80	0.00	0.00	0.00

Table 9.2 Membership Grade of Trade Volume Difference

Date	Trade Volume Difference (hundred million)	L_{21}	L_{22}	L_{23}	L_{24}	L_{25}
2003/1/3	325	0.00	0.00	0.00	0.20	0.80
2003/1/6	-56	0.00	0.62	0.38	0.00	0.00
2003/1/7	221	0.00	0.00	0.00	0.50	0.50
2003/1/8	110	0.00	0.00	0.00	0.82	0.18
2003/1/9	24	0.00	0.00	0.38	0.63	0.00
2003/1/10	-89	0.03	0.97	0.00	0.00	0.00
2003/1/13	116	0.00	0.00	0.00	0.80	0.20
2003/1/14	92	0.00	0.00	0.00	0.87	0.13
2003/1/15	-154	0.24	0.76	0.00	0.00	0.00
2003/1/16	171	0.00	0.00	0.00	0.64	0.36

The daily weighted stock price index fluctuation and trade volume high/low difference from 2003/January/3 to 2003/March/11 are shown in Table 9.1 and 9.2. Suppose the maximum membership grade of some day is located at L_{1j} ($j=1,2,...,5$), its linguistic variable will be regarded as L_{1j} ($j=1,2,...,5$). Take 2003/January/3 as an example. The maximum membership grade is located at L_{14} and L_{25}. Thus, the weighted stock price index fluctuation of 2003/January/3 is L_{14} and the trade volume difference is L_{25}. Or we can call the fluctuation value of the day as "up" and the trade volume difference as "very high." The fuzzy relationship among data can be located, based on past fuzzy data, and furthermore, fuzzy-Markov-relative matrix is obtained as well.

$$R=\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

$$\begin{bmatrix}
0.38 & 0.62 & 0.80 & 0.10 & 0.00 & 0.38 & 0.26 & 0.80 & 0.78 & 0.00 \\
0.33 & 0.75 & 0.94 & 0.75 & 0.30 & 0.80 & 0.94 & 0.77 & 0.64 & 0.23 \\
0.52 & 0.59 & 0.95 & 0.94 & 0.74 & 0.48 & 0.86 & 0.63 & 0.74 & 0.36 \\
0.99 & 0.61 & 0.94 & 0.75 & 0.58 & 0.52 & 0.75 & 0.70 & 0.95 & 1.00 \\
0.39 & 0.61 & 0.58 & 0.35 & 0.12 & 0.25 & 0.83 & 0.38 & 0.58 & 0.13 \\
0.20 & 0.51 & 0.49 & 1.00 & 0.25 & 0.34 & 0.86 & 0.36 & 0.64 & 0.26 \\
0.97 & 1.00 & 0.80 & 0.86 & 0.58 & 0.66 & 0.68 & 0.95 & 0.95 & 0.80 \\
0.52 & 0.74 & 0.77 & 0.63 & 1.00 & 0.22 & 0.94 & 0.63 & 0.95 & 0.38 \\
0.39 & 0.67 & 0.90 & 0.63 & 0.50 & 0.77 & 1.00 & 0.89 & 0.80 & 0.18 \\
0.23 & 0.36 & 0.48 & 0.98 & 0.50 & 0.36 & 0.74 & 0.38 & 0.50 & 0.18
\end{bmatrix}$$

R_{11} is the fuzzy-Markov-relative matrix of the stock price fluctuation of (a certain day) and the stock price fluctuation of (the next day). R_{12} is the fuzzy-Markov-relative matrix of the stock price fluctuation of (a certain day) and the trade volume difference of (the next day). R_{21} is the fuzzy-Markov-relative matrix of the trade volume difference of (a certain day) and the stock price fluctuation of (the next day). R_{22} is the fuzzy-Markov-relative matrix of the trade volume difference of (a certain day) and the trade volume difference of (the next day). Thus, the multivariate first order of auto-regression model is $(FX_{1,t}, FX_{2,t}) = (FX_{1,t-1}, FX_{2,t-1})R$. $(FX_{1,t-1}, FX_{2,t-1})$ and $(FX_{1,t}, FX_{2,t})$ each indicates the membership function of the linguistic variable of multivariate fuzzy set of Taiwan weighted stock price index fluctuation and trade volume high/low difference in Day t-1 and Day t. At last, Table 9.3 and 9.4 present the membership functions output by the model and the membership functions converted from linguistic vector index functions.

Table 9.3 Membership functions output by the model regarding TAIEX and trade volume

Date	Membership functions output by the model
2003/1/6	(0.89, 0.96, 1.13, 1.46, 1.08) (0.88, 1.39, 1.04, 1.15, 0.83)
2003/1/7	(1.49, 1.22, 1.49, 1.37, 1.16) (1.14, 1.37, 1.31, 1.49, 1.49)
2003/1/8	(0.91, 1.09, 1.20, 1.20, 1.20) (0.98, 1.20, 1.13, 1.20, 0.54))
2003/1/9	(0.84, 1.22, 1.37, 1.07, 0.95) (1.22, 1.37, 1.27, 1.35, 0.63)
2003/1/10	(0.91, 1.16, 1.16, 1.16, 1.03) (1.10, 1.16, 1.16, 1.16, 0.73)
2003/1/13	(1.84, 1.58, 1.67, 1.62, 1.16) (1.18, 1.43, 1.65, 1.82, 1.67)
2003/1/14	(0.81, 1.25, 1.38, 1.05, 0.92) (1.19, 1.38, 1.22, 1.38, 0.60)
2003/1/15	(0.91, 1.26, 1.80, 1.55, 1.24) (1.25, 1.74, 1.50, 1.54, 0.54)
2003/1/16	(1.37, 1.37, 1.37, 1.37, 1.16) (1.18, 1.28, 1.37, 1.37, 1.37)

Table 4.4 Membership functions converted from linguistic vector index functions regarding TAIEX and trade volum

Date	Membership Functions After Transformation
2003/1/6	(0, 0, 0, 1, 0) (0, 1, 0, 0, 0)
2003/1/7	(1, 0, 1, 0, 0) (0, 0, 0, 1, 1)
2003/1/8	(0, 0, 0, 1, 1) (0, 1, 0, 1, 0)
2003/1/9	(0, 0, 1, 0, 0) (0, 1, 0, 0, 0)
2003/1/10	(0, 1, 1, 1, 0) (0, 1, 1, 1, 0)

Date	Membership Functions After Transformation
2003/1/13	(1, 0, 0, 0, 0) (0, 0, 0, 1, 0)
2003/1/14	(0, 0, 1, 0, 0) (0, 1, 0, 1, 0)
2003/1/15	(0, 0, 1, 0, 0) (0, 1, 0, 0, 0)
2003/1/16	(1, 1, 1, 1, 0) (0, 0, 1, 1, 1)

Comparisons and Analyses of Prediction Results

Since this study focuses mainly on the property tendency of time series, we make use of the fuzzy rule base discussed in section 9.2 to verify the converted membership functions and output the properties of the prediction results. Furthermore, we have introduced the meaning and definition of belief function in detailed in section 9.5. The properties of the prediction results and the belief functions are listed in Table 9.5 and 9.6.

Table 9.5 Comparison of fitness value for TAIEX fluctuation

Date	True Value	Predicted Value of Multivariate Fuzzy Time Series	Belief Function of the Predicted Value
2003/1/6	up	up	0.71
2003/1/7	unchanged	plunge	0.65
2003/1/8	soar	soar	0.47
2003/1/9	unchanged	unchanged	0.73
2003/1/10	up	unchanged	0.58
2003/1/13	soar	plunge	0.51
2003/1/14	unchanged	unchanged	0.74
2003/1/15	up	unchanged	0.75
2003/1/16	down	down	0.47
2003/1/17	down	unchanged	0.75
2003/1/20	up	up	0.79
2003/1/21	unchanged	plunge	0.50
2003/1/22	up	up	0.78
2003/1/23	up	plunge	0.50
2003/1/24	up	up	0.74

Date	True Value	Predicted Value of Multivariate Fuzzy Time Series	Belief Function of the Predicted Value
2003/1/27	up	down	0.42
2003/1/28	up	up	0.72
2003/2/6	plunge	plunge	0.54
2003/2/7	down	unchanged	0.83
2003/2/10	down	unchanged	0.75
2003/2/11	down	up	0.75
2003/2/12	unchanged	down	0.60
2003/2/13	plunge	unchanged	0.69
2003/2/14	unchanged	down	0.60
2003/2/17	soar	soar	0.55
2003/2/18	down	unchanged	0.76
2003/2/19	down	unchanged	0.76
2003/2/20	unchanged	unchanged	0.80
2003/2/21	unchanged	unchanged	0.71
2003/2/24	up	up	0.65
2003/2/25	plunge	plunge	0.66
2003/2/26	unchanged	unchanged	0.81
2003/2/27	unchanged	unchanged	0.69
2003/3/3	up	up	0.44
2003/3/4	down	unchanged	0.73
2003/3/5	down	soar	0.48
2003/3/6	unchanged	unchanged	0.78
2003/3/7	down	down	0.63
2003/3/10	down	unchanged	0.74
2003/3/11	down	down	0.66
Matching Average: 0.53			
Average Rank-Forecasting Accuracy: 0.81			

Table 9.6 Comparison of fitness value for TAIEX trade volume

Date	True Value (after realization)	Predicted Value of Multivariate Fuzzy Time Series	Belief Function of the Predicted Value
2003/1/6	low	low	0.70
2003/1/7	very high	very high	0.48
2003/1/8	high	medium	0.79
2003/1/9	high	low	0.68
2003/1/10	low	medium	0.60
2003/1/13	high	high	0.67
2003/1/14	high	medium	0.78
2003/1/15	low	low	0.73
2003/1/16	high	very high	0.44
2003/1/17	very low	very low	0.61
2003/1/20	low	low	0.75
2003/1/21	very high	very high	0.52
2003/1/22	low	low	0.77
2003/1/23	very high	very high	0.50
2003/1/24	low	low	0.69
2003/1/27	very low	medium	tie-undecided
2003/1/28	low	low	0.70
2003/2/6	high	high	0.68
2003/2/7	high	medium	0.80
2003/2/10	very low	medium	0.67
2003/2/11	high	low	0.73
2003/2/12	medium	low	0.61
2003/2/13	low	low	0.72
2003/2/14	medium	high	0.59
2003/2/17	high	high	0.73
2003/2/18	low	low	0.77
2003/2/19	high	high	0.61
2003/2/20	medium	low	0.74
2003/2/21	low	low	0.72
2003/2/24	low	high	0.64
2003/2/25	high	very high	0.48

Date	True Value (after realization)	Predicted Value of Multivariate Fuzzy Time Series	Belief Function of the Predicted Value
2003/2/26	medium	medium	0.79
2003/2/27	medium	low	0.73
2003/3/3	high	medium	0.65
2003/3/4	medium	low	0.67
2003/3/5	high	high	0.70
2003/3/6	low	low	0.74
2003/3/7	medium	high	0.65
2003/3/10	low	low	0.75
2003/3/11	high	high	0.63
Matching Average: 0.55			
Average Rank-Forecasting Accuracy: 0.86			

As shown in Table 9.5 and 9.6, the multivariate fuzzy time series forecast model established in this study has proved itself to be a very effective predicting device. Because the predictions are made through the five divided sections, the matching average should be about 0.20, regarding either stock price fluctuation or trade volume difference. However, if we made predictions through the model in this study, the matching averages of stock price fluctuation and trade volume difference are 0.53 and 0.55, and the average forecasting accuracy are 0.81 and 0.86 respectively. Belief function has also been helpful in deciding the amount of capital to be invested and controlling risks. It is clearly manifested that if the belief function is higher, the properties we predict are in conformity with the properties of true value. On the contrary, when the predicted properties and the true ones are distinct, the belief functions of the prediction are usually lower. As shown in the following Table 9.7, the authors present the true value properties as well as the predicted value properties of TAIEX fluctuation from March 12 to April 23, generated by the model we propose in this study.

Table 9.7 True value, predicted value and belief function of TAIEX fluctuation

Date	True Value	Predicted Value of Multivariate Fuzzy Time Series unchanged	Belief Function of the Predicted Value
2003/3/12	up	unchanged	0.79
2003/3/13	up	unchanged	0.67
2003/3/14	up	unchanged	0.74
2003/3/17	plunge	up	0.58
2003/3/18	soar	up	0.74
2003/3/19	down	up	0.72
2003/3/20	up	unchanged	0.53
2003/3/21	unchanged	unchanged	0.66
2003/3/24	unchanged	up	0.65
2003/3/25	down	up	0.60
2003/3/26	unchanged	unchanged	0.73
2003/3/27	unchanged	unchanged	0.70
2003/3/28	down	unchanged	0.73
2003/3/31	plunge	unchanged	0.73
2003/4/1	unchanged	unchanged	0.81
2003/4/2	down	up	0.59
2003/4/3	up	down	0.59
2003/4/4	soar	unchanged	0.74
2003/4/7	up	unchanged	0.73
2003/4/8	unchanged	unchanged	0.71
2003/4/9	unchanged	unchanged	0.54
2003/4/10	unchanged	up	0.61
2003/4/11	unchanged	unchanged	0.71
2003/4/12	down	up	0.61
2003/4/15	up	down	0.68
2003/4/16	up	soar	0.45
2003/4/17	down	up	0.73
2003/4/18	up	unchanged	0.74
2003/4/21	unchanged	up	0.73

Date	True Value	Predicted Value of Multivariate Fuzzy Time Series unchanged	Belief Function of the Predicted Value
2003/4/22	down	up	0.62
2003/4/23	unchanged	unchanged	0.66
Matching Average: 0.27			
Average Rank-Forecasting Accuracy: 0.72			

Table 9.8 True value, predicted value and belief function of TAIEX fluctuation

Date	True Value	Predicted Value of Multivariate Fuzzy Time Series unchanged	Belief Function of the Predicted Value
2003/4/8	unchanged	unchanged	0.71
2003/4/9	unchanged	unchanged	0.54
2003/4/10	unchanged	up	0.61
2003/4/11	unchanged	unchanged	0.71
2003/4/12	down	up	0.61
2003/4/15	up	down	0.68
Matching Average: 0.5			
Average Rank-Forecasting Accuracy: 0.79			

Owing to the *Second Gulf War* during March, 20 to May, 2, 2003, the first SARS case identified on March, 8 in Taiwan, the seal-off of Taipei Municipal Hoping Hospital on April, 24, and the fast spread of SARS, Taiwan stock market had suffered great impact. Nonetheless, under the severe market unrest, the matching average for the next 31 days, as shown in Table 4.7, manages to reach 0.27 and the average forecasting accuracy, 0.72. If we exclude the ten days prior and post Gulf War and the days after SARS outbreaks, we find that there are six days with matching average as high as 0.5 and average forecasting accuracy, 0.79 as shown in Table 4.8. This has proved that the multivariate fuzzy time series forecast model we establish is, to a certain extent, credible in predicting future conditions. The reason of some prediction values' failing to reach the true values is that we only take the maximum membership function into consideration and neglect the secondary membership functions in the conversion process.

To sum up, with the help of a reasonable forecast model and belief functions for evaluating the prediction results, we can make an appropriate investment strategy and won't feel at a loss for future investment and can reduce the risks as well.

9.6 Conclusion Remarks

In the scientific research and analysis, the uncertainty and fuzziness contained in statistical data is often the obstacle for traditional model construction. If we use quasi-accurate value for cause and effect analysis or quantitative measurement, it will result in bias of cause and effect, misleading of decision model, and enlarging difference between prediction and real situation. Manski (1990) has pointed out that as numerical data contains the risk of over-demand and over-interpretation, the adoption of fuzzy numbers can help us avoid such a risk. Thus, it is very important to carefully examine fuzziness and robustness of numerical data in investigating quantitative method in social sciences. However, for those hard to explain cognitive questions, we can more clearly express them through membership functions and fuzzy statistical analysis. Hence, qualitative measurement and fuzzy statistics should be an advanced way to better describe human thinking and feeling.

In this research, we propose the definition of multivariate fuzzy time series, fuzzy relations, and fuzzy Markov relation matrix etc, and further construct a multivariate fuzzy time series model as well as forecasting. It is worth to mention that based on the knowledge extraction and derivative method, we use a threshold function via building fuzzy rule base to reach a corresponding linguistic variables. Lastly, we build a good integrated process of modeling and use this process to construct a forecasting model for price limit and trading volume difference of Taiwan Weighted Stock Index. Using daily price limit and trading volume difference of weighted index from January 15 2000 to February 21 2000 as historical data, we also establish a appropriate multivariate fuzzy times series model. Furthermore, we use average accuracy of forecasting to illustrate the forecasting performances of multivariate fuzzy time series models, traditional ARIMA models and Neural Networks and find that multivariate fuzzy time series model has the best forecasting performance.

Finally, in spite of the robust forecasting performance for multivariate fuzzy time series modeling, there remain some problems for further studies. For example:

(i) There are many factors influence weighted index such as trading volume, exchange rates, interest rate, and so on. In this research, we only consider closing price and trading volume for constructing multivariate fuzzy time series model, one more variables could be included to make result more accurate.

(ii) For multivariate fuzzy time series, we need to stabilize data first and then proceed with advanced analysis so that more accurate result can be obtained. However, how to judge whether multivariate fuzzy time series are stationary or not? We could also consider whether they have seasonality.

(iii) In this research, we adopt five-ranking classification and transform time series data to fuzzy numbers through membership functions. In the contrary, seven-ranking classification used in social sciences could be tried in future studies. Though we realize that the finer the classification the more complicated the calculation, however, it should provide certain improvement on forecasting performance.

(iv) The improvement of membership function modeling techniques will also increase analysis and forecasting performance for time series data in the integrated process of modeling.

(v) To simplify the transformation process of fuzzy numbers, we only consider the greatest membership and omit other memberships. Maybe we can conduct further study for this heuristic issue.

Reference

[1] Artstein, Z. and Vitale, R. (1975) A strong law of large numbers for random compact sets. *Ann. Prob*. 3, 879-882.

[2] Bagnoli, C., and Smith, H. C. (1998) The theory of fuzzy logic and its application to real estate valuation, *Journal of Real Estate Research*, 16(2), 169-199.

[3] Baldwin, J., Lawry, J., and Martin, T. (1998) The application of generalized fuzzy rules to machine learning and automated knowledge discovery. *Inter. J. Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(5), 459-487.

[4] Baraldi, A. and Blonda, P. (1999) A survey of fuzzy clustering algorithms for pattern recognition. Part I and Part II. *IEEE Trans. Syst., Man, Cybern. Part B: CYBERN*. 29, 6, 778-785 and 786-801.

[5] Billingsley, P (1968) *Convergence of Probability Measures*. J. Wiley, New York.

[6] Calvo, T. and Mesiar, R. (2001) Generalized Median. *Fuzzy Sets and Systems*, *124*, 59-61.

[7] Carrano A. L., Taylor, J. B., Young, R E., Lemaster R. L. and Saloni, D. E. (2004) Fuzzy knowledge-based modeling and statistical regression in abrasive wood machining, *Forest Products Journal*, 54(5), 66-72.

[8] Casalino, G., Memola, F. and Minutolo, C. (2004) A model for evaluation of laser welding efficiency and quality using an artificial neural network and fuzzy logic, *Proceedings of the Institution of Mechanical Engineers*, 218(6), 641.

[9] Chen, S. and Wu, B. (2003) On optimal forecasting with soft computation for nonlinear time series. *Fuzzy Optimization and Decision Making*. 3, 215-228.

[10] Chen, S. M. and Hwang, J. R. (2000) Temperature prediction using fuzzy time series. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 30, No. 2, 263-275.

[11] Chiang, D., Chow, L. and Wang, Y. (2000) Mining time series data by a fuzzy linguistic summary system. *Fuzzy Sets and Systems*, 112, 419-432.

[12] Choquet, G. (1953/54) Theory of Capacities. *Ann. Inst. Fourier.* 5, 131-259

[13] Cressie, N. (1979) A central limit theorem for random sets. *Z.* Wahr. Verw. Gebiet. 49, 37-47.

[14] Cressie, N. and Laslett, G.M. (1987) Random set theory and problems of modeling. *SIAM Review*, 29, 557-574.

[15] Dembo, A. and Zeitouni, O. (1998) Large Deviations Techniques and Applications. Springer-Verlag, New York.

[16] Dempster, A, P. (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* 38, 325-339.

[17] Diamond, P. and Kloden, P.(1994) *Metric Spaces of Fuzzy Sets.* World Scientific, Singapore.

[18] Dorsey, D. W., and Coovert, M. D. (2003) Mathematical modeling of decision making: A soft and fuzzy approach to capturing hard decisions, *Human Factors*, 45(1), 117.

[19] Dozzi, M., Merzbadi, E. and Schmidt, V. (2001) Limit theorems for sums of random fuzzy sets. *J. Math. Anal. and Appl.* 259, 554-565.

[20] Dubois, D. and H. Prade (1991) Fuzzy sets in approximate reasoning, Part I: Inference with possibility distributions. *Fuzzy Sets and Systems*, 40, 143-202.

[21] Dupuis, P. and Ellis, R. (1997) *A Weak Convergence Approach to the Theory of Large Deviations*. J. Wiley, New York.

[22] Esogbue, A. O., Song, Q. (2003) On optimal defuzzification and learning algorithms: theory and applications, *Fuzzy Optimization and Decision Making*, 2(4), 283.

[23] Feng, D, and Nguyen (2004) On statistical inference with random sets. *Soft methodology and Random Information systems* (M. Lopez *et al,* Eds.) 77-84. Springer-Verlag, New York.

[24] Feng, D.J. and Fenf, D. (2004) On a statistical framework for estimation from random set observations. *J. Theoretical Prob.* 17, 85-110.

[25] Frechet, M. (1948) éléments aléatoires de nature quclcnique, dans un espace distancié. *Ann. H. Paircaré, IV,* 215-310.

[26] Fuller, R. and Keresztvalvi, T. (1990) On generalizations of Nguyen's theorem. *Fuzzy Stes and Systems.* 4, 371-374.

[27] Galvo, T., and Mesiar, R. (2001) Generalized Medians, *Fuzzy Sets and Systems,* 124, 59-64.

[28] Gill, R. D., Van der Laan, M. J. and Robin, J. M. (1997) Coarsening at random characterizations, conjectures, counter examples. *Lectures Notes in Statistics.* 123, 255-294. Springer-Verlag, New York.

[29] Goodman, I, R. (1982) Fuzzy sets as equivalence classes of random sets. *Fuzzy Sets and Possibility Theory*. (R. Yager, Ed.), Pergamon, 327-343.

[30] Grabisch, M., Nguyen, H. and Walker, E. (1995) *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*. Kluwer Academic, Dordrecht.

[31] Guariso, G. , Rizzoli, A., and Werthner, H., Identification of model structure via qualitive simulation, *IEEE Trans. on Systems, Man, and cybernetics,* Vol. 22, No. 5, 1075-1086, 1992.

[32] Hajek, K. (1981) *Sampling from a finite population.* Marcell-Dekker, N.Y.

[33] Hartigan . J. A. (1987) Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* 82, 267-270,

[34] Heithan, D.F. and Rubin, D. B. (1991) Probability and coarse data. *Ann. Statis.* 19, 2244-2253.

[35] Foundations *On Fuzzy Set Theory*. Boston: Kluwr Academic.

[36] Huarng K. (2001) Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets and Systems*, 123, 369-386.

[37] Huber, P. and Strassen, V. (1973) Minimax tests and the Neyman-pearson leman for capacities. *Ann. Statist.* 1,251-263.

[38] Huber, P.J. (1973) The use of Choquet capacities in statistics. *Bull. Inst. Intern. Statist* (JLV, Book 4), 181-188.

[39] Kao, C., and Chyu, C. L. (2003) Least-squares estimates in fuzzy regression analysis, *European Journal of Operational Research*, 148(2), 426.

[40] Klein, E. and Thompson, A.C. (1984) *Theory of Correspondences* (including Applications to Mathematical Economics) John Wiley, New York.

[41] Korner, R. (1997) On the Variance of Fuzzy Random Variables. *Fuzzy Sets and Systems*, Vol. 92, p83-93.

[42] Kreinovich, V. Nguyen, H. and Wu, B. (1999) Justification of heuristic methods in data processing using fuzzy theory, with applications to detection of business cycles from fuzzy data. *East-West Journal of Mathematics.* **1**(2), 147-157.

[43] Kumar, K. and Wu, B. (2001) Detection of change points in time series analysis with fuzzy statistics. *International Journal of Systems Science*. **32**(9), 1185- 1192.

[44] Li, B. and Wang, T. (2005) On consistency of estimators based on random set observations. *J. Information Sciences* .To appear.

[45] Li, S, Ogura, Y. and Kreinovich, V. (2002) *Limit Theorems and Application of Set-valued and Fuzzy Set-valued Random Variables*. Kluwer Academic, Dardrecht.

[46] Lingras, P., and West, C. (2004) Interval Set Clustering of Web Users with Rough K-Means, *Journal of Intelligent Information Systems*, 23(1), 5.

[47] Liu, W. Y., and Song, N. (2001) The fuzzy association degree in semantic data models, *Fuzzy Sets and Systems*, 117, 203-208.

[48] Manski, C. (1990) The use of intention data to predict behavior: A Best Case Analysis. *Journal of the American Statistical Association*, 85, 934-940.

[49] Manton¸ K. Woodbury¸ M. and Tolley, H. (1994) *Statistical Applications – Using Fuzzy Set.* John Willy & Sons, Inc.¸ New York.

[50] Marinacci, M. and Montrucchio, L. (2004) *Uncertainty in Economics Theory* (I. Gilboa, Ed), Routledge, New York. 46-107.

[51] Matheran, G. (1975) *Random sets and Integral geometry*, J. Wiley. New York.

[52] Meyerowitz, A., Richman, F. and Walker, E. A. (1994) Calculating maximum entropy densities for belief functions. *Intern. J. Uncertainty, Fuzziness and Knowledge-Based Systems* 2(4), 377-389.

[53] Molchanov, I. (1993) *Limit Theorems for Unions of Random Closed Sets.* Lecture Notes in Mathematics. 1561. Springer-Verlag, New York.

[54] Molchanov, I. (2005) *Theory of Random Sets*. Springer-Verlag, Berlin.

[55] Mosteller, F. and Youtz, C. (1990) Quantifying probabilistic expressions. *Statis. Sci*. 5, 2-34.

[56] Nelsen, R. B. (1999) An Introduction to Copulas. *Lecture Notes in Statistics*, 139, Springer-Verlag, New York.

[57] Nguyen, H and Wu, B. (2000) *Fuzzy Mathematics and Statistical Applications.* Hua-Tai Book Company, Taipei

[58] Nguyen, H, Wang, T. and Wu, B. (2004) On probabilistic methods in fuzzy theory. *International Journal of Intelligent Systems.* 19, 99-109.

[59] Nguyen, H. and Kreinovich, V. (1999) How to divide a territory ? A new simple differential formalism for optimization of set functions. *J. Intelligent Systems.* 14, 223-251.

[60] Nguyen, H. and Wu, B. (1999) Fuzzy/Probability ~ Fractal/Smooth. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. **7**(4), 363-370.

[61] Nguyen, H. T (2006) *An introduction to Random sets*. Chapman and Hall/CRC. To appear.

[62] Nguyen, H. T. and Walker E. A. (2005) *A First Course in Fuzzy Logic* (Third Edition), Chapman and Hall/CRC. Boca Raton, Florida.

[63] Nguyen, H. T. Prasad, N. R., Walker C. and Walker E. (2003) *A First Course in Fuzzy and Neural Control*. Chapman and Hall/CRC. Boca Raton, Florida.

[64] Nguyen, H., Wu, B. and Kreinovich, V. (2000) On combining statistical and fuzzy techniques: detection of business cycles from uncertain data. in *Trends in Information Technology*, **1**, 69-74.

[65] Noberg, T. (1984) Convergence and existence of random set distributions with applications. *Israel J. Math.* 77, 241-264.

[66] Noberg, T. (1992) On the existence of ordered couplings of random set, with applications. *Israel J. Math.* 77, No. 3, 241-264.

[67] Orlowski, S. A. (1991) *Calculus of Decomposible Properties, Fuzzy Sets and Decisions*. Allertion Press.

[68] Park, Y. M., U. C. Moon, and K. Y. Lee (1995) A self-organizing fuzzy logic controller for dynamic systems using fuzzy auto-regressive moving average (FARMA) model. *IEEE Transactions on Fuzzy Systems*, Vol. 3, No. 1, 75-82.

[69] Peters, G. (2001) A linear forecasting model and its application to economic data, *Journal of Forecasting*, 20(5), 315.

[70] Polonick, W. (1995) Density estimation under qualitative assumptions in higher dimensions. *J. Mult. Anal.* 55, 61-81.

[71] Ross,T. (1995) *Fuzzy Logic with Engineering Applications.* McGraw-Hill, Inc., New York.

[72] Rubin, D. B. (1976) Inference and missing data. *Biometrika.* 63, 581-592.

[73] Ruspini, E. (1991) Approximate reasoning: past, present, future. *Information Sciences,* 57, 297-317.

[74] Saaty, T. L. (1977) A scaling method for priorities in hierarchical structures. *J. Math. Psychology.* 15, 234-281.

[75] Saaty, T. L. (1994) *Fundamentals of Decision Making and Priority Theory with the analytic Hierarchy Process*. PWS publications.

[76] Sabinetti, G. and Wets, R. (1986) On the convergence in distribution of Measurable multifunctions (random sets), normal integrands., Statistic processes and stochastic inference. *Math. Oper. Research*. 11, 385-419.

[77] Salinetti, G. and Wets, R. (1981) On the convergence of closed-valued measurable mutifunctions. *Trans. Amer. Math, Soc*. (1), 275-289.

[78] Sanchez J.D, and Gomez A.T. (2004) Estimating a fuzzy term structure of interest rates using fuzzy regression techniques, *European Journal of Operational Research*, 154 (3), 804-818.

[79] Sanchez J. D., and Gomez A.T. (2003) Estimating a term structure of interest rates for fuzzy financial pricing by using fuzzy regression methods, *Fuzzy Sets and Systems*, 139 (2), 313-331.

[80] Sato, M. and Sato, Y. (1994) On a multicriteria fuzzy clustering method for 3-way data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2, 127-142.

[81] Schreiber, T. (2000) Statistical inference from set-valued observations. *Prob. Math. Statist.* 20, 223-235.

[82] Schweizer, B. and Sklar, A. (1983) *Probabilistic Metric Spaces*. North-Holland, Armstersdam.

[83] Shapley, L. S. (1971) Cores of convex games. *Intern. J. Game Theory,* 1, 11-26.

[84] Singpurwalla, N.D. and Booker, J.M. (2004) Membership functions and probability measures of fuzzy sets. *J. Amer. Statist, Assoc.* 99, 867-889.

[85] Song, Q. and B. S. Chissom (1993b) Fuzzy time series and its models. *Fuzzy Sets and Systems*, 54, 269-277.

[86] Song, Q. and B. S. Chissom (1994) Forecasting enrollments with fuzzy time series—Part II. *Fuzzy Sets and Systems*, 62, 1-8.

[87] Stojakovic, M. (1994) Fuzzy Random Variables, Expectation, and Martingales . *Journal of Mathematical Analysis and Applications*, Vol. 184, p594-606.

[88] Sugeno, M. and K. Tanaka (1991) Successive identification of a fuzzy model and its applications to prediction of a complex system, *Fuzzy Sets and Systems*, 42, 315-334.

[89] Sullivan, J., and Woodall, W. H. (1994) Comparison of fuzzy forecasting and Markov modeling, *Fuzzy Sets and Systems*, 64, p279-293.

[90] Swapan, R., and Ray, K. (1997) Reasoning with vague default. *Fuzzy Sets and Systems*, 91, p327–338 .

[91] Syau Y. R., Hsieh, H. T., and Lee, E. S. (2001) Fuzzy Numbers in the Credit Rating of Enterprise Financial Condition, *Review of Quantitative Finance and Accounting*, 17(4), 351

[92] Tanaka, H., and Ishibuchi, H. (1993) Architecture of neural networks with interval weights and its application to fuzzy regression analysis. *Fuzzy Sets and Systems*, 57, 27-39.

[93] Tseng F. M., Tzeng G. H., and Yu H. C. (1999) Fuzzy seasonal time series for forecasting the production value of the mechanical industry in Taiwan. *Technological Forecasting and Social Change*, 60 (3), 263-273.

[94] Tseng F.M., Tzeng G. H., Yu H. C., and Yuan, J. C. (2001) Fuzzy ARIMA model for forecasting the foreign exchange market, *Fuzzy Sets and Systems*, 118, 9-19.

[95] Tseng, T. and C. Klein (1992) A new algorithm for fuzzy multicriteria decision making. *International Journal of Approximate Reasoning*, 6, 45-66.

[96] Turnquist, M. A., and Nozick, L. K. (2004) A nonlinear optimization model of time and resource allocation decisions in projects with uncertainty, *Engineering Management Journal*, 16(1), 40.

[97] Wallster, T. S. (1986) Measuring the vague meaning of probability terms. *J. Exp. Psychology.* 115, 348-365.

[98] Walters, P. (1982) *An Introduction to Ergodic Theory.* Springer-Verlag, New York.

[99] Wang H. F., and Tsaur, R. C. (2000) Resolution of fuzzy regression model, *European Journal of Operational Research*, 126(3), 637.

[100] Wang, C.H. (2004) Predicting tourism demand using fuzzy time series and hybrid grey theory. *Tourism Management*, 25(3), 367-374

[101] Wang, Z. and G. J. Klir (1992) *Fuzzy Measure Theory*. New York: Plenum Press.

[102] Watanabe, N. and Imaizumi, T. (1993) A fuzzy statistical test of fuzzy hypotheses. *Fuzzy Sets and Systems*, 53, 167-178.

[103] Wei C.P., Liu O. R., and Hu J.H. (2002) Fuzzy Statistics Estimation in Supporting Multidatabase Query Optimization, *Electronic Commerce Research*, 2(3),2 87. *287*

[104] Wen, F., Chen, L and Wu, B. (2000) Spatial time series analysis in 1997-1998 Eastern Asia financial crisis. *Asia Pacific Management Review*. **5**, 331-345.

[105] Wong, Z. and G.J. Klir (1992) *Fuzzy Measure Theory*. New York: Plenum Press.

[106] Wu, B. (1995) Fuzzy statistical analysis: a new research method on the sampling survey. *Research Letters of National Chengchi University.*, **2**, 65-80.

[107] Wu, B. and Hsu, Yu-Yun. (2004) A New approach of bivariate fuzzy time series: with applications to the stock index forecasting. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems.* 11(6), 671-690.

[108] Wu, B. and Hung, S. (1999) A fuzzy identification procedure for nonlinear time series: with example on ARCH and bilinear models. *Fuzzy Set and System.* 108, 275-287.

[109] Wu, B. and Chen, M. (1999) Use of fuzzy statistical technique in change periods detection of nonlinear time series. *Applied Mathematics and Computation,* 99,

[110] Wu, B. and Sun, C. (1996) Fuzzy statistics and computation on the lexical semantics. *Language, Information and Computation* (*PACLIC 11*), 337-346. *Seoul, Korea.*

[111] Wu, B. and Sun, C. (2001) Interval-valued statistics, fuzzy logic, and their use in computational semantics. *Journal of Intelligent and Fuzzy Systems.* **11**,1-7.

[112] Wu, B. and Tseng, N. (2002) A new approach to fuzzy regression models with application to business cycle analysis. *Fuzzy Sets and System*. 130, 33-42.

[113] Wu, B. and Hwang, Z. (1995) Testing for the stability of Taiwan's short run money demand function: the application of fuzzy time series. *Annual Conference Proceedings of Taiwan Economic Association.* 169-190.

[114] Wu, B. and Yang, W. (1997) Application of fuzzy statistics in the sampling survey. *Development and Application for the Quantity Methods of Social Science.* 289-316, Academic Sinica, Taiwan.

[115] Xu, B., Dale, D. S., Huang, Y., and Watson, M. D. (2002) Cotton color classification by fuzzy logic, *Textile Research Journal*, 72(6), 504-509.

[116] Yoshinari, Y., W. Pedrycz, and K. Hirota (1993) Construction of fuzzy models through clustering techniques. *Fuzzy Sets and Systems*, 54, 157-165.

[117] Zadeh, L. A. (1965) Fuzzy Sets. *Information and Control*, *8*, 338-353.

[118] Zadeh, L. A. (1968) Probability measures of fuzzy events. *J. Math. Anal and Appl,* 23, 421-427.

[119] Zadeh, L. A. (1971) Quantitative fuzzy semantics. *Information Sciences,* 3, 159-176.

[120] Zadeh, L. A. (1975) The Concept of linguistic variable and its application to approximate reasoning, parts I, II. *Information Sciences,* 8, 119-249; 8, 301-35.

[121] Zadeh, L. A. (2002) Toward a perception-based probabilities. *J. Statist. Planning and Inference*. 105, 233-264.

[122] Zimmermann, H. J. (1991) *Fuzzy Set Theory and Its Applications.* Boston: Kluwer Academic.

Index