MODELING UNCERTAINTY USING PROBABILISTIC BASED POSSIBILITY THEORY WITH APPLICATIONS TO OPTIMIZATION

by

Kenneth David Jamison

M.S., University of California at Riverside, 1981

B.S., University of California at Riverside, 1978

A thesis submitted to the

University of Colorado at Denver

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Applied Mathematics

1998

This thesis for the Doctor of Philosophy

degree by

Kenneth David Jamison

has been approved

by

Weldon Lodwick

Stephen Billups

Jan Mandel

Burt Simon

William Wolf

Date

Jamison, Kenneth David (Ph.D., Applied Mathematics)

Modeling Uncertainty Using Probabilistic Based Possibility Theory with Applications to Optimization

Thesis directed by Professor Weldon Lodwick

ABSTRACT

It is shown that possibility distributions can be formulated within the context of probability theory and that membership values of fuzzy set theory can be interpreted as cumulative probabilities. The basic functions and operations of possibility theory are interpreted within this setting. The probabilistic information that can be derived from possibility distributions is examined. This leads to two functionals that provide estimates for the expected value of a random variable, the expected average of a single possibility distribution and the estimated expectation that requires two special possibility distributions to compute. Secondly, the space of fuzzy numbers is examined. It is shown that this space can be partitioned into a vector space and that the expected average functional motivates a norm on this space. It is shown that for most applications, Cauchy sequences converge in this space. Thirdly, applications of this theory to problems in optimization are examined. The concept of a fuzzy function is formulated. The minimum and minimizer of a fuzzy function are described. Objectives in fuzzy optimization are examined. Lastly, the ideas of the thesis are applied to the linear programming problem with uncertain coefficients.

This abstract accurately represents the content of the candidate's thesis. I recommend its publication.

Signed ______________________

Weldon Lodwick

ACKNOWLEDGMENTS

I wish to thank Weldon Lodwick for his direction, support and encouragement, the fuzzy set theory study group (Weldon, Steve Russell, Chris Mehl and Ram Shanmugan) for their inspiration, Burt Simon for his challenging comments and Steve Billups for his very thoughtful review of this thesis. Most of all I thank my wife Liza O'Neill. Without her love and support I would not have been able to persevere.

CONTENTS

1. Introduction and Overview

Suppose a decision maker (DM) is interested in optimizing (in some sense) the function

$$f(x) = a^2x - b \tag{1.1}$$

where a and b are fixed but unknown (uncertain) parameters. How does the DM model the unknown parameters a and b? Given the model, how does the DM evaluate and interpret f(x)=a$^2x - b$? How is a decision reached? Two common techniques for estimating the parameters a and b are interval analysis and probability theory. In this thesis we will consider a third technique, called possibility theory.

The advantage of interval analysis is that the interval solution is guaranteed to contain all solutions (see Moore [32]). Interval analysis also offers the advantage of being computationally tractable. It is often straightforward to evaluate the function using the known methods of interval arithmetic (Moore [32]). For example, if a$\in [1,3]$ and b$\in [4,6]$ then

$$f(1) \in [1,3]^2 - [4,6] = [1,9] - [4,6] = [-5,5]. \tag{1.2}$$

The evaluation gives an interval which represents the range of possible values of the function, though most often this range is overestimated (see Moore [32]). The DM is then faced with the problem of optimizing an interval valued function, i.e. the DM must decide on an ordering of the set of intervals. One approach is to assume a uniform probability distribution over any given interval and optimize the expected value, i.e. optimize the midpoint of the interval. For example, f(2)$\in$[1,3]$^2 2 - [4, 6]$ =[2,18]-[4,6]=[-4,14] would be considered greater than f(1)$\in$ $[-5, 5]$ since the midpoint (14-4)/2=5 is greater than the midpoint (5-5)/2=0. A disadvantage of interval analysis is that it gives the worst case and interval solutions can be large (useless) if care and computationally intensive approaches are not used. Another disadvantage is that it gives no consideration to the likelihood of a particular outcome. It only considers the range of all possible outcomes.

Probability theory considers not only the range of possible outcomes but the likelihood a particular outcome may occur. For example, suppose a and b are modeled as the independent random variables X and Y with the following probability density functions:

$$f_X(x) = \begin{cases} x - 1 \text{ for } x \in [1, 2] \\ \\ 3 - x \text{ for } x \in [2, 3] \end{cases} \quad \text{and } f_Y(y) = \begin{cases} \text{y}-4 \text{ for y} \in [4, 5] \\ \\ 6 - y \text{ for y} \in [5, 6] \end{cases} . \qquad (1.3)$$

Then the probability density function for the random variable Z=f(1)=X$^2 - Y$

is the following:

$f_Z(z) =$

$\int_1^{\sqrt{6+z}}(w-1)(6-w^2+z)dw$	z∈[-5,-4]
$\int_{\sqrt{5+z}}^{\sqrt{6+z}}(w-1)(6-w^2+z)dw + \int_1^{\sqrt{5+z}}(w-1)(w^2-z-4)dw$	z∈[-4,-3]
$\int_{\sqrt{5+z}}^{\sqrt{6+z}}(w-1)(6-w^2+z)dw + \int_{\sqrt{4+z}}^{\sqrt{5+z}}(w-1)(w^2-z-4)dw$	z∈[-3,-2]
$\int_{\sqrt{5+z}}^{2}(w-1)(6-w^2+z)dw + \int_2^{\sqrt{6+z}}(3-w)(6-w^2+z)dw$ $+ \int_{\sqrt{4+z}}^{\sqrt{5+z}}(w-1)(w^2-z-4)dw$	z∈[-2,-1]
$\int_{\sqrt{5+z}}^{\sqrt{6+z}}(3-w)(6-w^2+z)dw + \int_{\sqrt{4+z}}^{2}(w-1)((w^2-z)-4)dw$ $+ \int_2^{\sqrt{5+z}}(3-w)(w^2-z-4)dw$	z∈[-1,0]
$\int_{\sqrt{5+z}}^{\sqrt{6+z}}(3-w)(6-w^2+z)dw + \int_{\sqrt{4+z}}^{\sqrt{5+z}}(3-w)(w^2-z-4)dw$	z∈[0,3]
$\int_{\sqrt{5+z}}^{3}(3-w)(6-w^2+z)dw + \int_{\sqrt{4+z}}^{\sqrt{5+z}}(3-w)(w^2-z-4)dw$	z∈[3,4]
$\int_{\sqrt{4+z}}^{3}(3-w)(w^2-z-4)dw$	z∈[4,5]

Using this approach, the DM might seek to optimize the expected value of f(x). For example, the expected value of Z=f(1), E(Z), is -.8333. A disadvantage of this approach is that it may be impossible, difficult and/or impractical to obtain the probability distribution for the random variable f(x) for each x to be considered in the optimization problem. In addition, some problems do not admit to a probabilistic formulation (e.g. locating a tumor for radiation therapy).

In this thesis we examine a third approach called possibility theory.

It is argued that possibility theory is a compromise between interval analysis and probability theory. In essence, we perform interval arithmetic over families of intervals where the families of intervals are parameterized by a probability distribution. The result is a family of intervals representing the possible values of f(x) that preserves some of the characteristics of the probability distribution. There are two principle advantages to be gained from this approach. One advantage is that problems formulated in possibilitistic terms are more computationally tractable. Another advantage is that possibility theory allows for computations with imprecise probabilities. For example, a possibility distribution can be formulated from a family of confidence intervals. These advantages will become clear in the sequel.

Consider the example above where a and b were modeled as independent random variables X and Y with density functions given in (1.3). We will show that two parameterized families of intervals can be constructed for X and Y to arrive at the following:

$$X_\alpha^L = \begin{cases} [1, 2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right) + 1] \text{ for } \alpha \in [.75, 1] \\ [1, 2\left(1 - \frac{1}{2}\sqrt{\left(2 - 2\sqrt{(-\alpha+1)}\right)}\right) + 1] \text{ for } \alpha \in [0, .75] \end{cases}$$

and

$$Y_\alpha^L = \begin{cases} \left[4, 2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right) + 4\right] \text{ for } \alpha \in [.75, 1] \\ \left[4, 2\left(1 - \frac{1}{2}\sqrt{\left(2 - 2\sqrt{(-\alpha+1)}\right)}\right) + 4\right] \text{ for } \alpha \in [0, .75] \end{cases} .$$

and

$$X_\alpha^R = \begin{cases} [3-2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right), 3] \text{ for } \alpha \in [.75, 1] \\ [3-2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right), 3] \text{ for } \alpha \in [0, .75] \end{cases}$$

and

$$Y_\alpha^R = \begin{cases} \left[6-2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right), 6\right] \text{ for } \alpha \in [.75, 1] \\ \left[6-2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right), 6\right] \text{ for } \alpha \in [0, .75] \end{cases} .$$

Then two parameterized families of intervals can be constructed for Z=f(1) using interval arithmetic as follows:

$$Z_\alpha^R =$$
$$\begin{cases} [\left(3-2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)^2 - \left(2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}+4\right), 5] \text{ for } \alpha \in [.75, 1] \\ \left[\left(3-2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)\right)^2 \right. \\ \qquad \left. - \left(2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)+4\right), 5\right] \text{ for } \alpha \in [0, .75] \end{cases}$$

and

$$Z_\alpha^L$$
$$= \begin{cases} [-5, \left(2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)+1\right)^2 - \left(6-2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)] \text{ for } \alpha \in [.75, 1] \\ \left[-5, \left(2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right)+1\right)^2 \right. \\ \qquad \left. - \left(6-2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)\right)\right] \text{ for } \alpha \in [0, .75] \end{cases}$$

The DM must now rank these families of intervals in order to reach a decision. We will show that in general these two families of intervals provide an upper and a lower bound for the cumulative probability distribution function of $Z = f(1)$. If the goal of the DM is to optimize the expected value, this characterization allows for the calculation of an estimated expected value (in this case an actual value). For this example, the calculation of the estimated expected value of Z, which we denote as EE(Z), is as follows:

$$
\begin{aligned}
&\mathrm{EE(Z)} \\
&= \frac{1}{2}[\int_{.75}^{1}\left(\left(3-2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)^{2}-\left(2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}+4\right)\right)d\alpha \\
&+ \int_{0}^{.75}\left(\left(3-2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)\right)^{2}\right. \\
&- \left.\left(2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)+4\right)\right)d\alpha \\
&+ \int_{.75}^{1}\left(\left(2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)+1\right)^{2}-\left(6-2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)\right)d\alpha \\
&+ \int_{0}^{.75}\left(\left(2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right)+1\right)^{2}\right. \\
&- \left.\left(6-2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)\right)\right)d\alpha] \\
&= -0.83333.
\end{aligned}
$$

The focus of this thesis is to explore the application of possibility theory to the problem of incorporating uncertainty into math modeling, particularly problems in optimization. The thesis will examine the use of possibility distributions as a method of modeling unknown parameters in real valued functions. The contributions of this work to the mathematical foundation and application of possibility theory include (i) a formulation of possibility theory within the framework of probability theory (ii) a method for selecting among families of possibility distributions over the real line based on an estimate of the expected value (iii) a norm on the space of fuzzy numbers based on this estimated expected value (iv) an isometry between this normed spaced and the subspace of $L^1[0,1]$ (Lebesgue integrable functions on the interval [0,1]) consisting of functions of bounded variation (v) convergence properties of this normed space (vi) a new definition of a fuzzy function and several implications of this definition and (vii) an approach to linear programming problems where the coefficients are stated in terms of fuzzy numbers. Many of these results are a compilation and continuation of research that will appear in publications Jamison&Lodwick [16], Jamison [17] and Jamison&Lodwick [15]. Additional aspects of this research have appeared in Lodwick&Jamison [28] and Lodwick&Jamison [29].

1.1 Overview

The second chapter of this paper develops topics in possibility theory as used here. The chapter begins with known definitions and basic implications from possibility theory. Then it is shown that a possibility distribution can be interpreted as a cumulative probability distribution. Several results are presented showing how the image of a possibility distribution under an arbitrary function is defined and how the resulting possibility distribution is interpreted.

Chapter three focuses on possibility distributions over the real line. Of particular interest are fuzzy numbers, i.e. unknown numbers characterized by possibility distributions. There are possibilistic distributions over linguistic variables but these are not examined here. The mathematics of fuzzy numbers is presented. It is shown how fuzzy numbers can be added, subtracted, multiplied, etc. An equivalence relation on the space of fuzzy numbers is presented and it is shown that this results in a vector space. A norm on this space is defined. This norm is motivated from an estimated expected value functional. The convergence properties in this normed space are examined.

In chapter four, applications to certain optimization problems are examined. First, the concept of a fuzzy function between finite dimensional vector spaces is discussed. A fuzzy function is defined as a possibility distribution, satisfying certain properties, over the space of bounded functions. The image of a fuzzy function is shown to be a fuzzy vector. In particular, if the range space is the real line, the image is a fuzzy number.

Second, the case of minimizing an unconstrained fuzzy real-valued convex function is examined. The minimum of a fuzzy function is defined and it is shown to be a fuzzy number. The minimizer of a fuzzy function is defined. It is shown that the possibility distribution for the minimizer has a connectedness property. A minimization approach is examined based on the estimated expected value functional of chapter two. It is called the method of minimum regrets, and seeks to minimize the estimated expected value of the possible error.

In chapter five, the theory and methods of chapters two, three and four are applied to the linear programming problem where all of the coefficients are replaced by fuzzy numbers. The problem is first formulated in possibilistic terms and converted to a problem of maximizing the estimated expected value of an unconstrained fuzzy function. It is shown that the resulting problem is concave. Conditions to insure a bounded solution are developed and an upper bound on the decision variables is determined. An algorithm is provided.

Chapter six concludes by summarizing the results of this thesis and suggesting areas of future research.

2. Possibility Theory

L. Zadeh [48] introduced the concept of a fuzzy set in the 1960's. The basic idea is to extend the notion of set membership to include degrees of partial membership. This is generally characterised by a **membership function** which is a mapping from the universe of discourse into the unit interval. For example, if A is a fuzzy subset of set X, the membership function for A is $\mu_A : X \to [0,1]$. For x$\in X$, $\mu_A(x)$ is called the **membership value** of x and is interpreted as the degree to which x is a member of fuzzy set A where 1 represents full membership and 0 represents complete lack of membership. If $\mu_A(x) = 0$ or 1 for all x, A is called a **crisp set** (in this case μ_A is just the characteristic function of set A). Zadeh defined a set of rules of combination for fuzzy sets, each of which is a generalization of the same concepts for crisp sets. So, for example, the standard union, intersection and complement operations for fuzzy sets A and B, defined in terms of their membership functions are $\mu_{A\cup B}(x) = \max(\mu_A(x), \mu_B(x))$, $\mu_{A\cap B}(x) = \min(\mu_A(x), \mu_B(x))$, and $\mu_{A^c}(x) = 1 - \mu_A(x)$ where A^c is the complement of A. Note that these are the usual operations if A and B are crisp sets.

Possibility theory, also introduced by L. Zadeh [49], is a natural extension of his work with fuzzy sets. As a tool in optimization problems that involve uncertainty, the basic idea is to consider the set of all possible alternatives for an unknown parameter of the problem. But the set of possible alternatives may not be well defined. Thus, a fuzzy set of alternatives is used. To do this, each element in the universal set is assigned a membership value (or **possibility level**) in the interval [0,1]. In possibility theory, as formulated by Zaden, the membership value of an element is interpreted as the degree of possibility that the parameter is that element. If the membership value is zero, it is impossible that the parameter is the element. If the membership value is one the element is considered a possible value of the parameter without reservation. Membership values produce an ordering on the universal set in terms of each elements acceptability as a possible value of the unknown parameter. The mapping of each member of the universal set to its membership value is called a **possibility distribution function** for the unknown parameter.

The typical process for utilizing possibility theory in optimizing a function with ill-defined parameters is as follows: First, each unknown parameter of the function is represented using a possibility distribution. Second, for a given point in the decision space the function is evaluated over the set of possible parameter values using the rules of combination of possibility theory.

This produces a possibility distribution for the function evaluation. Third, the possibility distribution is assigned a value. The process of assigning a value to a possibility distribution is called **defuzzification** (taken from fuzzy set theory). Fourth, this value is used to order the possibility distributions so that an optimal decision can be made. It is the second of these steps that makes possibility theory an interesting alternative to probability theory in modeling uncertainty. For many problems, it is fairly straight forward to calculate the image of a function of possibility distributions, where it may be very difficult if probability distributions are used.

The process just described is controversial. The basic problem is that membership values have not been well defined except as a tool for ordering the universal set. Very different membership values can be used to arrive at the same ordering. But most defuzzification methods are dependent on the membership values. Thus the decision derived from the process may be arbitrary, since it depends on what membership values were assigned. Also, it is difficult to interpret the meaning of the number derived from defuzzification. In [12], Dubois&Prade state "...when we scan the fuzzy set literature, including Zadeh's own papers, there is no uniformity in the interpretation of what a membership grade means. This situation has caused many a critique by fuzzy set opponents, and also many a misunderstanding within the field itself. Most negative statements expressed in the literature turn around the question of interpreting and eliciting membership grades."

This thesis begins with an interpretation of membership values based on probability theory. The fundamental properties of possibility theory are then examined from this perspective. The advantage of this approach is that it gives a concrete meaning to membership values and the value derived from the defuzzification method proposed.

2.1 Possibility Measures and Possibility Distribution Functions

Before introducing the probabilistic interpretation of membership values, the formal definitions and resulting implications from possibility theory are provided. Note that the basis for selecting membership values is not discussed in this section.

Definition 1 (Klir&Yuan [21]) Given a universal set X and a nonempty family C of subsets of X, a **fuzzy measure** on $\langle X, C\rangle$ is a function

$$g : C \rightarrow [0, 1] \tag{2.1}$$

that satisfies the following requirements:

($g1$) g($\emptyset$) = 0 and g(X)=1 (boundary condition)

(g2) $\forall$ A,B$\in C$, if A$\subseteq B$ then g(A)$\leq g(B)$ (monotonicity) (g3) for any increasing sequence $A_1 \subset A_2 \subset ...$ in C, if $\bigcup_{i=1}^{\infty} A_i \in C$, then

$$\lim_{i \to \infty} g(A_i) = g(\bigcup_{i=1}^{\infty} A_i) \tag{2.2}$$

(continuity from below)

(g4) for any decreasing sequence $A_1 \supset A_2 \supset ...$ in C, if $\bigcap_{i=1}^{\infty} A_i \in C$, then

$$\lim_{i \to \infty} g(A_i) = g(\bigcap_{i=1}^{\infty} A_i) \tag{2.3}$$

(continuity from above).

An immediate consequence of this definition is that if $A, B, A \cup B$ and $A \cap B \in C$, then every fuzzy measure satisfies the inequalities,

$$g(A \bigcap B) \leq \min(g(A), g(B)) \text{ and } g(A \bigcup B) \geq \max(g(A), g(B)). \tag{2.4}$$

Fuzzy measures include probability measures as well as belief and plausibility measures of evidence theory (see Klir&Yuan [21]). Possibility theory is based on the following two fuzzy measures.

Definition 2 (Klir&Yuan [21]) Let nec denote a fuzzy measure of $\langle X, C \rangle$. Then nec is called a **necessity measure** iff

$$nec\left(\bigcap_{k \in K} A_k\right) = \inf_{k \in K} nec\,(A_k) \tag{2.5}$$

for any family $\{A_k \mid k \in K\}$ in C such that $\bigcap_{k \in K} A_k \in C$, where K is an arbitrary index set.

Definition 3 (Klir&Yuan [21]) Let pos denote a fuzzy measure of $\langle X, C\rangle$. Then pos is called a **possibility measure** iff

$$pos\left(\bigcup_{k\in K} A_k\right) = \sup_{k\in K} pos\,(A_k) \tag{2.6}$$

for any family $\{A_k \mid k \in K\}$ in C such that $\bigcup_{k\in K} A_k \in C$, where K is an arbitrary index set.

When $C = \mathcal{P}(X)$, the power set of X, possibility and necessity measures occur in pairs.

Theorem 1 (Klir&Yuan [21]) If pos is a possibility measure on $\mathcal{P}(X)$. Then its dual set function nec, which is defined by

$$nec(A) = 1 - pos(A^c) \tag{2.7}$$

is a necessity measure called the necessity measure associated with pos.

Definition 4 (Klir&Yuan [21]) Given a possibility measure pos on power set $\mathcal{P}$(X), the function $\mu : X \to [0,1]$ such that $\mu(x) = pos\,(\{x\})$ for all x$\in X$ is called the **possibility distribution function** associated with pos and $\mu(x)$ is called the **possibility level** or **membership value** of x.

Theorem 2 (Klir&Yuan [21]) Every possibility measure pos on a power set $\mathcal{P}$(X) is uniquely determined, for each $A \in \mathcal{P}(X)$, by the associated possibility distribution function via the formula

$$pos(A) = sup_{x\in A}\mu(x). \tag{2.8}$$

Theorem 3 (Wang&Klir [46]) If μ is a possibility distribution function for possibility measure pos then

$$sup_{x \in X}\mu(x) = 1. \tag{2.9}$$

Conversely, if a function $\mu : X \rightarrow [0, 1]$ satisfies (2.9), then μ can determine a possibility measure pos uniquely, and μ is the possibility distribution function of pos.

Definition 5 (Dubois&Prade [10]) A possibility distribution is called **normal** if $\exists a \in X$ such that pos({a})=1.

2.2 Membership Values as Cumulative Probabilities

Given a possibility distribution function, there are various ways membership values have been interpreted. For example, membership values have been interpreted as probabilities over nested sets as in Dubois&Prade [11] or as upper bounds on probability distributions as in Dubois&Prade [9]. A good overview of the various interpretations of membership grades in fuzzy set theory as well as possibility theory can be found in Dubois&Prade [12]. In this thesis we adopt an interpretation that is closely related to the random set interpretation and the interpretation as the upper bounds on probability distributions. This interpretation provides a straight forward method for constructing possibility distributions. This interpretation will give context to the operations proposed in this thesis. The need for a concrete interpretation of membership values was discussed in the introduction to this chapter.

The prototype for the following discussion is a parameterized real valued function where the parameters are not known with precision. However, other applications of possibility theory will be considered.

Definition 6 Let X be a space and $\tilde{x}$ a random vector on X. A **possibility nest** for $\tilde{x}$ is a relation $\geq_p$ on $X \times X$ such that

(1) $\geq_p$ is a linear ordering of X with the property that $\forall\ x, y \in X$ either $x\ >_p\ y$, $y\ >_p\ x$ or $x =_p y$ and

(2) $\forall x \in X$, $\{y \in X \mid y \geq_p x\}$ is measurable with respect to the probability measure for $\tilde{x}$ and

(3) $\forall B \subseteq X$, $\exists$ an at most countable subset C $\subseteq$B with the property that $\forall x \in B\ \exists x_n, x_m \in C$ such that $\mathrm{x}_n \leq_p x \leq_p x_m$.

This third property requires some examination. The assumption is not trivial since there exists an unbounded infinite well ordered set that has the property that every countable subset is bounded above (see the set S_Ω in Munkres [33]). However, the following theorem shows that the assumption is not too restrictive.

Theorem 4 Let $B \subseteq [0, 1]$. Then $\exists$ an at most countable subset $C \subseteq B$ with the property that $\forall x \in B\ \exists x_n, x_m \in C$ such that $x_n \geq x \geq x_m$.

Proof:

We first show that there is a countable dense subset of B. For each $q \in Q \cap [0,1]$ (where Q is the set of rational numbers). Let $S(q, \frac{1}{n}) = \left\{x \in [0,1] \mid |x-q| < \frac{1}{n}\right\}$. Let $b_{q,n} \in S(q, \frac{1}{n}) \cap B$ if $S(q, \frac{1}{n}) \cap B \neq \emptyset$ otherwise set $b_{q,n} = b'$, an arbitrary element of B. Let $D = \bigcup_{q \in Q \cap [0,1]} \{b_{q,n} \mid n = 1, ...\}$. Then D is at most countable since it is a countable union of countable sets. Let $b \in B$. Pick $n > 0$. Since the rationals are dense in $[0,1]$ $\exists q$ rational such that $|q-b| < \frac{1}{2n}$. Then $b \in S(q, \frac{1}{2n})$ so $S(q, \frac{1}{2n}) \cap B \neq \emptyset$ which means $\exists b_{q,2n}$ in D such that $|q - b_{q,2n}| < \frac{1}{2n}$. But then $|b - b_{q,2n}| < |b-q| + |q - b_{q,2n}| = \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}$. Hence D is dense in B.

Let $C = D \cup (\sup D \cap B) \cup (\inf D \cap B)$. We claim C has the desired property of the theorem. Assume not. Then without loss of generality, assume $\exists b \in B$ such that $\forall c \in C$, $c < b$. Since D is dense in B $\exists \{d_n \in D \mid n = 1, ...\}$ such that $d_n \to b$. But then by assumption on b, $b = \sup D$ thus $b \in C$.$\square$

We motivate the naming possibility nest by noting that any such ordering produces a nested family of measurable subsets of X.

Definition 7 Let $\geq_p$ be a possibility nest for $\tilde{x}$. Define x to be **possible** if x$\geq_p \tilde{x}$ and **impossible** otherwise.

The following theorem provides some of the motivation behind these definitions.

Theorem 5 Let $\Delta = \{A_\alpha \mid \alpha \in \Lambda\}$ be a nested family of confidence intervals for random variable $\tilde{x}$ where $\alpha = prob(\tilde{x} \in A_\alpha)$. Then Δ defines a possibility nest for $\tilde{x}$.

Proof:

Define an ordering on R by setting $x >_p y$ if $\sup\{1-\alpha | x \in A_\alpha\} > \sup\{1-\alpha | y \in A_\alpha\}$ where $\sup\emptyset \equiv 0$. This clearly establishes a linear ordering on R since every two elements are comparable and the relation is transitive. For given $x \in R$ let $1-\gamma = \sup\{1-\alpha | x \in A_\alpha\}$. We claim that

$$\{y | y \geq_p x\} = \begin{cases} \cap_{\alpha\in\Lambda|\alpha>\gamma} A_\alpha & \text{if } \inf\{\alpha \in \Lambda | \alpha > \gamma\} = \gamma \\ \cap_{\alpha\in\Lambda|\alpha>\gamma} A_\alpha & \text{otherwise.} \end{cases}$$

$(\subseteq)$

Let $z = \{y | y \geq_p x\}$ then $z \geq_p x$ so $\sup\{1-\alpha | z \in A_\alpha\} \geq 1-\gamma$. Since the sets are neseted and indexed by a probability measure it must hold that $\alpha > \beta \Rightarrow A_\alpha \supset A_\beta$ or equivalently $1-\alpha < 1-\beta \Rightarrow A_\alpha \supset A_\beta$. Therefore $\sup\{1-\alpha | z \in A_\alpha\} \geq 1-\gamma$ means $z \in A_\alpha \; \forall \alpha > \gamma$ since otherwise $\exists \beta > \gamma$ for which $z \notin A_\beta$ but then $z \notin A_\alpha \; \forall \alpha \leq \beta$ and hence $\sup\{1-\alpha | z \in A_\alpha\} \leq 1-\beta < 1-\gamma$ which is a contradiction. Therefore, in all cases $z \in \cap_{\alpha\in\Gamma|\alpha>\gamma} A_\alpha$.

Assume $\inf\{\alpha \in \Lambda | \alpha > \gamma\} = \beta > \gamma$. If $\gamma \in \Lambda$ and if $z \notin A_\gamma$ then $z \notin A_\alpha \; \forall \alpha < \beta$ and $\sup\{1-\alpha | z \in A_\alpha\} \leq 1-\beta < 1-\gamma$ which is a contradiction. Therefore $z \in \cap_{\alpha\in\Lambda|\alpha\geq\gamma} A_\alpha$. If $\gamma \notin \Lambda$ then $\cap_{\alpha\in\Lambda|\alpha\geq\gamma} A_\alpha = \cap_{\alpha\in\Lambda|\alpha>\gamma} A_\alpha$ and the prior case holds.

$(\supseteq)$

Assume $\inf\{\alpha \in \Lambda | \alpha > \gamma\} = \gamma$. Let $z \in \cap_{\alpha>\gamma} A_\alpha$. Then $\sup\{1-\alpha | z \in A_\alpha\} \geq 1-\gamma$ and $z \in \{y | y \geq_p x\}$.

Assume $\inf\{\alpha \in \Lambda | \alpha > \gamma\} = \beta > \gamma$. If $z \in \cap_{\alpha \in \Lambda | \alpha \geq \gamma} A_\alpha$ then in particular if $\gamma \in \Lambda$ then $z \in A_\gamma$ so $\sup\{1-\alpha | z \in A_\alpha\} \geq 1-\gamma$ and $z \in \{y | y \geq_p x\}$. But it must be that $\gamma \in \Lambda$ since otherwise $1-\gamma = \sup\{1-\alpha | x \in A_\alpha\}$ would imply that $\inf\{\alpha \in \Lambda | \alpha > \gamma\} = \gamma$ contradicting our assumption.

The third property of the definition of possibility nest follows from the prior theorem and the fact that $\Lambda \subseteq [0,1]$. It remains to show that $\cap_{\alpha \in \Lambda | \alpha > \gamma} A_\alpha$ and $\cap_{\alpha \in \Lambda | \alpha \geq \gamma} A_\alpha$ are measurable. But by assumption, $\exists$ at most countable subsets C and D of Λ such that $\cap_{\alpha \in \Lambda | \alpha > \gamma} A_\alpha = \cap_{c \in C} A_c$ and $\cap_{\alpha \in \Lambda | \alpha \geq \gamma} A_\alpha = \cap_{d \in D} A_d$ and both of these latter sets are measurable since each A_α is measureable by assumption. $\square$

We are interested in functions of multiple random vectors. Therefore, we extend the notion of a possibility nest to a finite collection of random vectors as follows:

Definition 8 Let X be a space and $\widetilde{X} = \{\tilde{x}^i \mid i = 1 \text{ to N}\}$ be a finite set of symbols representing random vectors on X. Let $\Psi = \widetilde{X} \times X$. A **possibility nest in the context of** $\widetilde{X}$ is defined to be a relation, $\geq_c$, on $\Psi \times \Psi$ such that

(1) $\geq_c$ is an linear ordering of Ψ with the property that $\forall(\widetilde{x}^i, a), (\widetilde{x}^j, b) \in \Psi$ either $(\widetilde{x}^i, a) >_c (\widetilde{x}^j, b)$, $(\widetilde{x}^j, b) >_c (\widetilde{x}^i, a)$ or $(\widetilde{x}^i, a) =_c (\widetilde{x}^j, b)$ and

(2) $\forall$ a $\in X$ and $1 \leq i, j \leq N$, $\{b \in X \mid (\widetilde{x}^i, b) \geq (\widetilde{x}^j, a)\}$ is measurable with respect to the probability measure associated with random vector $\widetilde{x}^i$ and

(3) $\forall B \subseteq \Psi$, $\exists$ an at most countable subset C $\subseteq$B with the property that $\forall$ $(\widetilde{x}^i, a) \in B$, $\exists(\widetilde{x}^j, a_n), (\widetilde{x}^k, a_m) \in C$ such that $(\widetilde{x}^j, a_n) \leq_c (\widetilde{x}^i, a) \leq_c (\widetilde{x}^k, a_m)$.

A **possibility nest for $\widetilde{x}^i$ in the context of $\widetilde{X}$** is defined to be the relation $\geq_p$ on $X \times X$ which is the restriction of $\geq_c$ to X by setting $b \geq_p a$ if $(\widetilde{x}^i, b) \geq_c (\widetilde{x}^i, a)$. Note that the properties of $\geq_c$ ensure that $\geq_p$ satisfies the properties of Definition 6.

In a little while we will construct contextual possibility distributions from contexual possibility nests. One can think of contextual possibility distributions as interim distributions on the way to constructing a possibility distribution for a function of the random vectors in $\widetilde{X}$. Examples of orderings on Ψ and its usefulness are given throughout the rest of this chapter.

From this point forward, we will use $(X, \widetilde{X}, \geq_c)$ or $(X, \widetilde{x}, \geq_p)$ to denote possibility nests. Note that $(X, \widetilde{X}, \geq_c)$ includes the special case $(X, \widetilde{x}, \geq_p)$. Ψ will always be used to represent $\widetilde{X} \times X$.

For each i, the symbol $\widetilde{x}^i$ represents an unknown element of X. For some unique $X_i \in X$, $\widetilde{x}^i = X_i$, i.e. X_i is the realization of $\widetilde{x}^i$. For purposes of the following definition we will distinguish between the symbol for the random variable, $\tilde{x}^i$ and the actual value of the random variable X_i. We will drop this distinction after this definition.

Definition 9 Given $\left(X, \widetilde{X}, \geq_c\right)$, let $\omega = \min\{(\tilde{x}^i, X_i) \mid i = 1 \text{ to N}\}$ where min is with respect to the ordering $\geq_c$. Given $(\tilde{x}^i, a) \in \Psi$, call $\tilde{x}^i = a$ **possible in the context of** $\widetilde{X}$ if $(\tilde{x}^i, a) \geq_c \omega$ and **impossible in the context of** $\widetilde{X}$ otherwise.

When $\widetilde{X}$ consists of a single variable, $(\tilde{x}^1, X_1) = \omega$ then a is possible in the context of $\widetilde{X}$ iff $a \geq_p X_1$, i.e. a is possible. In this case the ordering is a possibilty nest for $\tilde{x}$ and the reference to the context is dropped.

Given $\left(X, \widetilde{X}, \geq_c\right)$, let

$$\Psi_\gamma = \{\psi \in \Psi \mid \psi >_c \gamma\} \tag{2.10}$$

and let

$$\pi_i(\Psi_\gamma) = \left\{a \in X \mid (\tilde{x}^i, a) \in \Psi_\gamma\right\}. \tag{2.11}$$

For $\gamma \in \Psi$, if $\gamma \geq_c \omega$ then for some i, $\tilde{x}^i \notin \pi_i(\Psi_\gamma)$. Recall that $\omega = \min\{(\tilde{x}^i, x_i)|1 = 1, \cdots, N\}$. We will show ω has the cumulative probability distribution characterized by:

$$\begin{aligned} F_\omega(\gamma) = 1 - prob(\tilde{x}^1 \in \pi_1(\Psi_\gamma) \mid \tilde{x}^2 \in \pi_2(\Psi_\gamma), \\ \cdots, \quad \tilde{x}^N \in \pi_N(\Psi_\gamma)) \cdots prob(\tilde{x}^N \in \pi_N(\Psi_\gamma)), \end{aligned} \tag{2.12}$$

where $F_\omega(\gamma) = prob(\gamma \geq_c \omega)$. If the $\widetilde{x}^i$'s are independent this reduces to

$$F_\omega(\gamma) = 1 - prob(\widetilde{x}^1 \in \pi_1(\Psi_\gamma)) \cdots prob(\widetilde{x}^N \in \pi_N(\Psi_\gamma)). \tag{2.13}$$

Thus $F_\omega(\gamma)$ is the probability that γ is possible in the context of $\widetilde{X}$. Then the statement "the probability is α that $\widetilde{x}^i = a$ is possible in the context of $\widetilde{X}$" has this specific meaning, namely that $prob((\widetilde{x}^i, a) \geq_c \omega) = \alpha$.

Theorem 6 Given $\left(X, \widetilde{X}, \geq_c\right)$, ω is a random element of Ψ and F_ω is the cumulative distribution function for ω relative to the ordering $\geq_c$.

Proof:

The $\sigma - field$ of subsets and probability measure follow from the definition of possibility nest in the context of $\widetilde{X}$ and F_ω is clearly the distribution function.□

Note that if $(\widetilde{x}^i, r) \geq_c \omega$ then this is also true for any $(\widetilde{x}^j, s) \geq_c (\widetilde{x}^i, r)$. For this situation,

$$prob\left((\widetilde{x}^i, r) \geq_c \omega \text{ and } (\widetilde{x}^j, s) \geq_c \omega\right) = prob\left((\widetilde{x}^i, r) \geq_c \omega\right) \tag{2.14}$$

and

$$prob\left((\widetilde{x}^i, r) \geq_c \omega \text{ or } (\widetilde{x}^j, s) \geq_c \omega\right) = prob\left((\widetilde{x}^j, s) \geq_c \omega\right). \tag{2.15}$$

Furthermore,

$$prob\left((\widetilde{x}^i, r) <_c \omega \text{ and } (\widetilde{x}^j, s) <_c \omega\right) \tag{2.16}$$

$$= \quad 1 - prob\left((\widetilde{x}^j, s) \geq_c \omega\right) = prob\left((\widetilde{x}^j, s) <_c \omega\right).$$

and finally,

$$\begin{aligned} & prob\left((\widetilde{x}^i, r) <_c \omega \text{ or } (\widetilde{x}^j, s) <_c \omega\right) && (2.17) \\ = \quad & 1 - prob\left((\widetilde{x}^i, r) \geq_c \omega\right) = prob\left((\widetilde{x}^i, r) <_c \omega\right). \end{aligned}$$

Note also that $(\widetilde{x}^j, s) \geq_c (\widetilde{x}^i, r)$ implies that

$$prob\left((\widetilde{x}^j, s) \geq_c \omega\right) \geq prob\left((\widetilde{x}^i, r) \geq_c \omega\right) \tag{2.18}$$

and

$$prob\left((\widetilde{x}^j, s) <_c \omega\right) \leq prob\left((\widetilde{x}^i, r) <_c \omega\right). \tag{2.19}$$

These observations generalize as follows.

Theorem 7 Given $\left(X, \widetilde{X}, \geq_c\right)$, the following identities hold $\forall$ B$\subset$ Ψ with B$\neq \emptyset$,

$$prob(\bigvee_{(\widetilde{x}^i, r) \in B} (\widetilde{x}^i, r) \geq_c \omega) = \sup\left\{prob((\widetilde{x}^i, r) \geq_c \omega) \mid (\widetilde{x}^i, r) \in B\right\} \tag{2.20}$$

and

$$prob(\bigwedge_{(\widetilde{x}^i, r) \in B} (\widetilde{x}^i, r) \geq_c \omega) = \inf\left\{prob((\widetilde{x}^i, r) \geq_c \omega) \mid (\widetilde{x}^i, r) \in B\right\} \tag{2.21}$$

and

$$prob(\bigwedge_{(\widetilde{x}^i,r)\in B} (\widetilde{x}^i,r) \ <\ _c\omega) = \inf\left\{prob((\widetilde{x}^i,r) <_c \omega) \mid (\widetilde{x}^i,r) \in B\right\} \quad (2.22)$$
$$= \ 1 - \sup\left\{prob((\widetilde{x}^i,r) \geq_c \omega) \mid (\widetilde{x}^i,r) \in B\right\}$$

and

$$prob(\bigvee_{(\widetilde{x}^i,r)\in B} (\widetilde{x}^i,r) \ <\ _c\omega) = \sup\left\{prob((\widetilde{x}^i,r) <_c \omega) \mid (\widetilde{x}^i,r) \in B\right\} \quad (2.23)$$
$$= \ 1 - \inf\left\{prob((\widetilde{x}^i,r) \geq_c \omega) \mid (\widetilde{x}^i,r) \in B\right\}$$

Proof:

For the first identity, note that

$$\left(\bigvee_{(\widetilde{x}^i,r)\in B} (\widetilde{x}^i,r) \geq_c \omega\right) = \bigcup_{(\widetilde{x}^i,r)\in B} \left((\widetilde{x}^i,r) \geq_c \omega\right)$$

and the events $((\widetilde{x}^i,r) \geq_c \omega)$ are nested because of the linear ordering $\geq_c$. By definition, $\exists$ an at most countable subset C $\subseteq$B with the property that $\forall(\widetilde{x}^i,a) \in B\ \exists(\widetilde{x}^j,a_n),(\widetilde{x}^k,a_m) \in C$ such that $(\widetilde{x}^j,a_n) \leq_c (\widetilde{x}^i,a) \leq_c (\widetilde{x}^k,a_m)$. This with the nested property imply that

$$\bigcup_{(\widetilde{x}^i,r)\in B} \left((\widetilde{x}^i,r) \geq_c \omega\right) = \bigcup_{n\in C} \left((\widetilde{x}^i,r_n) \geq_c \omega\right)$$

and

$$prob\left(\omega \in \bigcup_{(\widetilde{x}^i,r)\in B} ((\widetilde{x}^i,r) \geq_c \omega)\right) \ = \sup_{n\in C} \left\{prob\left((\widetilde{x}^i,r_n) \geq_c \omega\right)\right\}$$
$$= \sup_{(\widetilde{x}^i,r)\in B} \left\{prob\left((\widetilde{x}^i,r) \geq_c \omega\right)\right\}.$$

The other identies are similarly proven along with the fact that

$$inf\ \{prob((\tilde{x}^i, r) <_c \omega) \mid (\tilde{x}^i, r) \in B\}$$

$$= \inf \{1 - prob((\tilde{x}^i, r) \geq_c \omega) \mid (\tilde{x}^i, r) \in B\}$$

$$= 1 - \sup \{(\tilde{x}^i, r) \geq_c \omega) \mid (\tilde{x}^i, r) \in B\}$$

with a similar result using sup in place of inf.□

2.3 Possibility Distributions when Membership Values are Cumulative Probabilities

In this section it is shown that possibility distributions can be formulated from the cumulative probability distribution of ω as a random element of Ψ. This gives a probabilistic interpretation to membership values and to the possibility and necessity measures and other measures of possibility theory as will be shown.

Given $\left(X, \widetilde{X}, \geq_c\right)$, for $(\tilde{x}^i, a) \in \Psi$, define a function $\mu_{\tilde{x}^i}(a)$, by the equation

$$\mu_{\tilde{x}^i}(a) = prob((\tilde{x}^i, a) \geq_c \omega). \tag{2.24}$$

In other words, $\mu_{\tilde{x}^i}(a)$ is the probability that $\tilde{x}^i = a$ is possible in the context of $\widetilde{X}$ (or $\tilde{x} = a$ is possible, when $\widetilde{X} = \{\tilde{x}\}$).

Thus $\mu_{\tilde{x}^i}(a) = F_\omega((\tilde{x}^i, a))$ where $F_\omega((\tilde{x}^i, a))$ (see (2.12)) is the cumulative distribution function for the distribution of ω when ω is considered a random element of Ψ.

Theorem 8 Given $\left(X, \widetilde{X}, \geq_c\right)$, for $\widetilde{x}^i \in \widetilde{X}$, the function $\mu_{\widetilde{x}^i}$ is a possibility distribution function.

Proof:

By definition, $\mu_{\widetilde{x}^i} : X \to [0,1]$. Also

$$sup_{x \in X} \mu_{\widetilde{x}^i}(x) = \sup_{x \in X} prob\left(\left(\widetilde{x}^i, x\right) \geq_c \omega\right).$$

By definition $\exists$ an at most countable set C $\subseteq X$ with the property that

$$\begin{aligned}
&\sup_{x \in X} prob\left(\left(\widetilde{x}^i, x\right) \geq_c \omega\right) \\
&= \sup_{x \in C} prob\left(\left(\widetilde{x}^i, x\right) \geq_c \omega\right) \\
&= prob\left(\bigcup_{x \in C} \left(\widetilde{x}^i, x\right) \geq_c \omega\right) \\
&= prob\left(\bigcup_{x \in X} \left(\widetilde{x}^i, x\right) \geq_c \omega\right) = 1
\end{aligned}$$

since $(\widetilde{x}^i, X_i) \geq_c \omega$ by definition.$\square$

With this interpretation of membership values, it is clear that possibility distributions are context dependent in the following sense. If additional random vectors are added to $\widetilde{X}$ and/or removed from $\widetilde{X}$, then the possibility distribution for $\widetilde{x}^i$ will change. It will also change if the ordering of Ψ is changed. In fact, for a given collection of random vectors, $\widetilde{X}$, there is a set of possibility distributions (one for each ordering of Ψ).

The next theorem is a construction theorem for the case where the random vectors in $\widetilde{X}$ are independent. It is straightforward to construct a possibility distribution for a random vector $\widetilde{x}$ if $\widetilde{X}$ contains only the singleton $\widetilde{x}$. In this case $\omega = (\widetilde{x}^1, X_1)$ and the possibility nest forms a nested sequence of subsets of X (the sets $\pi(\Psi_{(\widetilde{x},a)})$, see (2.11)). Then $\mu_{\widetilde{x}}(a) = 1 - prob(\widetilde{x} \in \pi(\Psi_{(\widetilde{x},a)})) = prob(a \geq_p \widetilde{x}^i)$. For example, in Theorem 5 we saw that a nested sequence of confidence intervals can be a possibility nest for a real valued random variable. Once possibility distributions have been constructed in this manner for a collection of independent random variables, the following theorem shows how the possibility distributions can be combined to produce contextual possibility distributions for each of the variables. By contextual we mean the possibility distributions produced by an ordering on Ψ when $\widetilde{X}$ is the set of random variables of interest.

Theorem 9 Let $\widetilde{X} = \{\widetilde{x}^i \mid i = 1 \text{ to N}\}$ be a collection of independent random vectors on space X. Let $\left\{\mu^i_{\widetilde{x}^i} \mid i = 1 \text{ to N}\right\}$ be possibility distribution functions constructed for each $\widetilde{x}^i$ alone. In other words, for each i, an ordering has been imposed on X such that $\mu^i_{\widetilde{x}^i}(a) = prob(a \geq_p \widetilde{x}^i)$ with respect to this ordering. Then an order, $\geq_c$ can be imposed on the $\Psi = \widetilde{X} \times X$ by setting $(\widetilde{x}^i, a) >_c (\widetilde{x}^j, b)$ if $\mu^i_{\widetilde{x}^i}(a) > \mu^j_{\widetilde{x}^j}(b)$ and $(\widetilde{x}^i, a) =_c (\widetilde{x}^j, b)$ if $\mu^i_{\widetilde{x}^i}(a) = \mu^j_{\widetilde{x}^j}(b)$. For fixed a$\in X$ and i$\in \{1, 2, ..., N\}$, and for each j=1,2,...,N let

$$\alpha_j = 1 - \sup\left\{\mu^j_{\tilde{x}^j}(b) \mid b \in X \text{ and } \mu^i_{\tilde{x}^i}(a) \geq \mu^j_{\tilde{x}^j}(b)\right\}. \tag{2.25}$$

Then with respect to $\geq_c$,

$$\mu_{\tilde{x}^i}(a) = 1 - \prod_{j=1}^{N} \alpha_j \tag{2.26}$$

is a possibility distribution function for $\tilde{x}^i$ in the context of $\widetilde{X}$.

Proof:

From (2.13)

$$\mu_{\tilde{x}^i}(a) = 1 - prob(\tilde{x}^1 \in \pi_1(\Psi_{(\tilde{x}^i,a)})) \cdots prob(\tilde{x}^N \in \pi_N(\Psi_{(\tilde{x}^i,a)}))$$

where $\Psi_\gamma = \{\psi \in \Psi \mid \psi >_c \gamma\}$ and $\pi_i(\Psi_\gamma) = \{a \in X \mid (\tilde{x}^i, a) \in \Psi_\gamma\}$. But from the order imposed,

$$\Psi_{(\tilde{x}^i,a)} = \left\{(\tilde{x}^j, b) \mid \mu^j_{\tilde{x}^j}(b) > \mu^i_{\tilde{x}^i}(a)\right\} \text{ and } \pi_j(\Psi_{(\tilde{x}^i,a)}) = \left\{b \in X \mid \mu^j_{\tilde{x}^j}(b) > \mu^i_{\tilde{x}^i}(a)\right\}$$

and thus

$$\begin{aligned}
&\text{prob}(\tilde{x}^j \in \pi_j(\Psi_{(\tilde{x}^i,a)})) \\
&\quad = \; prob\left(\tilde{x}^j \in \left\{b \in X \mid \mu^j_{\tilde{x}^j}(b) > \mu^i_{\tilde{x}^i}(a)\right\}\right) \\
&\quad = \; 1 - prob\left(\tilde{x}^j \in \left\{b \in X \mid \mu^j_{\tilde{x}^j}(b) \leq \mu^i_{\tilde{x}^i}(a)\right\}\right).
\end{aligned}$$

But the event $(\widetilde{x}^j \in \left\{b \in X \mid \mu^j_{\widetilde{x}^j}(b) \le \mu^i_{\widetilde{x}^i}(a)\right\}) = \cup_{\mu^j_{\widetilde{x}^j}(b) \le \mu^i_{\widetilde{x}^i}(a)}(b \ge \widetilde{x}^j)$. Since these events are nested, and using property three of the defintion of a possibility nest, we have prob$\left(\cup_{\mu^j_{\widetilde{x}^j}(b) \le \mu^i_{\widetilde{x}^i}(a)}(b \ge \widetilde{x}^j)\right) = \sup\{prob(b \ge \widetilde{x}^j) | \mu^j_{\widetilde{x}^j}(b) \le \mu^i_{\widetilde{x}^i}(a)\}$. But prob$(b \ge \widetilde{x}^j) = \mu^j_{\widetilde{x}^j}(b)$. Therefore $1 - prob(\widetilde{x}^j \in \{b \in X | \mu^j_{\widetilde{x}^j}(b) \le \mu^i_{\widetilde{x}^i}(a)\}) = 1 - \sup\{\mu^j_{\widetilde{x}^j}(b) | b \in X | \mu^j_{\widetilde{x}^j}(b) \le \mu^i_{\widetilde{x}^i}(a)\}$. □

Corollary 1 If in addition to the other assumptions, the possibility distribution function $\mu^i_{\widetilde{x}^i}$ in Theorem 9 maps onto the interval (0,1) for each $i = 1$ to N then

$$\mu_{\widetilde{x}^i}(a) = 1 - \left(1 - \mu^i_{\widetilde{x}^i}(a)\right)^N \tag{2.27}$$

is a possibility distribution function for $\widetilde{x}^i$ in the context of $\widetilde{X}$.

Proof: Given $\mu^i_{\widetilde{x}^i}(a) \in (0,1)$, by assumption $\exists b \in X$ such that $\mu^j_{\widetilde{x}^j}(b) = \mu^i_{\widetilde{x}^i}(a)$. Therefore $\mu^i_{\widetilde{x}^i}(a) = \sup\left\{\mu^j_{\widetilde{x}^j}(b) \mid b \in X \text{ and } \mu^j_{\widetilde{x}^j}(b) \le \mu^i_{\widetilde{x}^i}(a)\right\}$. If $\mu^i_{\widetilde{x}^i}(a) = 1$ we know from the definition of a possibility distribution function that

$\sup\left\{\mu^j_{\widetilde{x}^j}(b) \mid b \in X \text{ and } \mu^j_{\widetilde{x}^j}(b) \le 1\right\} = 1$.

If $\mu^i_{\widetilde{x}^i}(a) = 0$, $\sup\left\{\mu^j_{\widetilde{x}^j}(b) \mid b \in X \text{ and } \mu^j_{\widetilde{x}^j}(b) \le 1\right\} = 0$ since either $\exists b$ such that $\mu^j_{\widetilde{x}^j}(b) = 0$ or $\left\{\mu^j_{\widetilde{x}^j}(b) \mid b \in X \text{ and } \mu^j_{\widetilde{x}^j}(b) \le 1\right\} = \emptyset$ and we defined $\sup \emptyset = 0$.

□

Let $F_i : R \to [0,1]$ be the cumulative distribution function for random variable $\widetilde{x}^i$ (i.e. $F_i(x) = prob\,(\widetilde{x}^i \leq x)$). In addition, let $G_i(x) = prob\,(x \leq \widetilde{x}^i)$. These functions map R into the interval [0,1] and satisfy the property that $\sup\{F_i(x) \mid x \in R\} = \sup\{G_i(x) \mid x \in R\} = 1$. Therefore these are possibility distribution functions for the case $\widetilde{X} = \{\widetilde{x}^i\}$ and the ordering on R is either the natural ordering or the reverse of the natural ordering. Then $\mu^i_{\widetilde{x}^i}(x) = prob\,(x \text{ is possible}) = prob\,(x \geq_p \widetilde{x}^i) = prob\,(x \geq \widetilde{x}^i) = F_i(x)$ or $\mu^i_{\widetilde{x}^i}(x) = prob\,(x \text{ is possible}) = prob\,(x \geq_p \widetilde{x}^i) = prob\,(x \leq \widetilde{x}^i) = G_i(x)$. When these functions satisfy the conditions of the previous corollary, they can be used directly to construct contextual possibility distributions. These special possibility distributions will be useful later when we examine the probabilistic information that can be derived from a possibility distribution.

Corollary 2 If $\widetilde{X}$ consists of random variables and for each element of $\widetilde{X}$ the distribution functions F_i and G_i map onto the interval $(0,1)$ (for example, if each $\widetilde{x}^i$ is a continuous random variable), then selection of one each of F_i or G_i induces an ordering of Ψ determined as in Theorem 9 and for each $i = 1$ to N

$$\mu_{\widetilde{x}^i}(a) = 1 - (1 - H_i(a))^N \tag{2.28}$$

is a possibility distribution for $\widetilde{x}^i$ in the context of $\widetilde{X}$ where $H_i = F_i$ or G_i.

Proof:

Apply the previous corollary. □

Note that when $\widetilde{X}=\{\tilde{x}\}$, this reduces to $\mu_{\tilde{x}}(a)=F(a)$ or $\mu_{\tilde{x}}(a)=G(a)$.

These distributions concentrate the possibility distributions to the left or right hand side of the real line.

Example 1 Constructing Possibility Distributions-Application of Theorem 9. In this example we will construct contextual possibility distribution functions. The first step is to construct possibility distribution functions for each random variable separately. Second, we use these distributions to construct the set $\Psi = \widetilde{X} \times X$ as in Theorem 9. Third, for each element of the set Ψ we construct the cumulative distribution function for ω, F_ω. Finally, using F_ω, the contextual possibility distribution functions for each element of $\widetilde{X}$ are constructed.

Let $\tilde{x}^1, \tilde{x}^2$ be independent discrete random variables over the set $X = \{1, 2, 3, 4, 5\}$ where the distributions of $\tilde{x}^1$ and $\tilde{x}^2$ are:

r	P($\tilde{x}^1 = r$)	P($\tilde{x}^2 = r$)
1	.25	.3
2	.5	.5
3	.125	.2
4	.125	0
5	0	0

We will construct two possibility distributions in the context of $\widetilde{X} = \{\tilde{x}^1, \tilde{x}^2\}$, one concentrating possibility to the left and one to the right side of the real line. We will denote these using superscripts L and R. Construct possibility distributions for each variable separately. Recall that $\mu(x) = prob((\tilde{x}, x) \geq_c \omega)$ which reduces to $prob(x \geq_p \tilde{x})$ when $\widetilde{X}$ consists of the single variable $\tilde{x}$. For $\tilde{x}^1$, the ordering for the left distribution is $1 >_p 2 >_p 3 >_p 4 >_p 5$ and for the right distribution, $4 >_p 3 >_p 2 >_p 1 >_p 5$. For $\tilde{x}^2$, the ordering for the left distribution is $1 >_p 2 >_p 3 >_p 4 =_p 5$ and for the right distribution, $3 >_p 2 >_p 1 >_p 4 =_p 5$.

r	${}^L\mu^1_{\tilde{x}^1}(r)$	${}^L\mu^2_{\tilde{x}^2}(r)$
1	1	1
2	.75	.7
3	.25	.2
4	.125	0
5	0	0

r	${}^R\mu^1_{\tilde{x}^1}(r)$	${}^R\mu^2_{\tilde{x}^2}(r)$
1	.25	.3
2	.75	.8
3	.875	1
4	1	0
5	0	0

We can combine one distribution for each of the random variables. We will combine ${}^L\mu^1_{\tilde{x}^1}(r)$ with ${}^L\mu^2_{\tilde{x}^2}(r)$ and ${}^R\mu^1_{\tilde{x}^1}(r)$ with ${}^R\mu^2_{\tilde{x}^2}(r)$ to get two sets of possibility distributions. As in Theorem 9, we construct an order on $\Psi = \{\tilde{x}^1, \tilde{x}^2\} \times X$, by setting $(\tilde{x}^i, a) >_c (\tilde{x}^j, b)$ if $\mu^i_{\tilde{x}^i}(a) > \mu^j_{\tilde{x}^j}(b)$ and set $\mu_{\tilde{x}^i}(a) = 1 - \prod_{j=1}^{N} \alpha_j$ where $\alpha_j = 1 - \sup\left\{\mu^j_{\tilde{x}^j}(b) \mid b \in X \text{ and } \mu^i_{\tilde{x}^i}(a) \geq \mu^j_{\tilde{x}^j}(b)\right\}$. The result is shown below with the elements of Ψ listed from highest to lowest in the ordering along with prob($\gamma \geq_c \omega$) $= 1 - prob(\gamma <_c \omega)$:

Ψ for ${}^L\mu^1_{\tilde{x}^1}(r)$ with ${}^L\mu^2_{\tilde{x}^2}(r)$	1-prob($(\tilde{x}^i, x) <_c \omega$)
$(\tilde{x}^1, 1) =_c (\tilde{x}^2, 1)$	$1 - (1-1)(1-1) = 1.0$
$(\tilde{x}^1, 2)$	$1 - (1-.75)(1-.7) = .925$
$(\tilde{x}^2, 2)$	$1 - (1-.25)(1-.7) = .775$
$(\tilde{x}^1, 3)$	$1 - (1-.25)(1-.2) = .4$
$(\tilde{x}^2, 3)$	$1 - (1-.125)(1-.2) = .3$
$(\tilde{x}^1, 4)$	$1 - (1-.125)(1-0) = .125$
$(\tilde{x}^1, 5) =_c (\tilde{x}^2, 4) =_c (\tilde{x}^2, 5)$	$1 - (1-0)(1-0) = 0$

Ψ for ${}^R\mu^1_{\tilde{x}^1}(r)$ with ${}^R\mu^2_{\tilde{x}^2}(r)$	1-prob($(\tilde{x}^i, x) < \omega$)
$(\tilde{x}^1, 4) = (\tilde{x}^2, 3)$	$1 - (1-1)(1-1) = 1.0$
$(\tilde{x}^1, 3)$	$1 - (1-.875)(1-.8) = .975$
$(\tilde{x}^2, 2)$	$1 - (1-.75)(1-.8) = .95$
$(\tilde{x}^1, 2)$	$1 - (1-.75)(1-.3) = .825$
$(\tilde{x}^2, 1)$	$1 - (1-.25)(1-.3) = .475$
$(\tilde{x}^1, 1)$	$1 - (1-.25)(1-0) = .25$
$(\tilde{x}^1, 5) = (\tilde{x}^2, 4) = (\tilde{x}^2, 5)$	$1 - (1-0)(1-0) = 0$

The later table gives us the following information:

P($\tilde{x}^1 = 4$ and $\tilde{x}^2 = 3$) $= 1 - .975 = .025$

P($\tilde{x}^1 = 3$ or 4 and $\tilde{x}^2 = 3$) $= 1 - .95 = .05$

P(($\widetilde{x}^1 = 3$ or 4 and $\widetilde{x}^2 = 2$ or 3) $= 1 - .825 = .175$

P($\widetilde{x}^1 = 2, 3,$ or 4 and $\widetilde{x}^2 = 2$ or 3)$=1 - .475 = .525$

P($\widetilde{x}^1 = 1, 2, 3$ or 4 and $\widetilde{x}^2 = 2$ or 3)$=1 - .25 = .75$

P($\widetilde{x}^1 = 1, 2, 3$ or 4 and $\widetilde{x}^2 = 1, 2$ or 3)$=1 - 0 = 1$

The left and right possibility distributions for $\widetilde{x}^1$ and $\widetilde{x}^2$ in the context of $\widetilde{X}$ are:

r	${}^L\mu_{\widetilde{x}^1}(r)$	${}^L\mu_{\widetilde{x}^2}(r)$
1	1	1
2	.925	.775
3	.4	.3
4	.125	0
5	0	0

r	${}^R\mu_{\widetilde{x}^1}(r)$	${}^R\mu_{\widetilde{x}^2}(r)$
1	.25	.475
2	.825	.95
3	.975	1
4	1	0
5	0	0

Example 2 The continuous Case - Application of Corollary 2 Let $\widetilde{x}^1$ and $\widetilde{x}^2$ be independent random variables with density functions

$$f_1(x) = \begin{cases} x - 1 \text{ for } x \in [1,2] \\ 3 - x \text{ for } x \in [2,3] \end{cases} \quad \text{and } f_2(x) = \begin{cases} x - 4 \text{ for } x \in [4,5] \\ 6 - x \text{ for } x \in [5,6] \end{cases} \tag{2.29}$$

so the cumulative distribution functions are:

$$F_1(x) = \begin{cases} \frac{1}{2}x^2 - x + \frac{1}{2} \text{ for } x \in [1,2] \\ 3x - \frac{1}{2}x^2 - 3.5 \text{ for } x \in [2,3] \end{cases}$$

$$\text{and} \quad F_2(x) = \begin{cases} \frac{1}{2}x^2 - 4x + 8 \text{ for } x \in [4,5] \\ \\ -17.0 + 6x - \frac{1}{2}x^2 \text{ for } x \in [5,6] \end{cases}$$

and the distribution functions from the left (where $G(x) = prob(x \leq \tilde{x})$ are:

$$G_1(x) = \begin{cases} .5 - \frac{1}{2}x^2 + x \text{ for } x \in [1,2] \\ \\ \frac{9}{2} - 3x + \frac{1}{2}x^2 \text{ for } x \in [2,3] \end{cases}$$

$$\text{and} \quad G_2(x) = \begin{cases} -7.0 - \frac{1}{2}x^2 + 4x \text{ for } x \in [4,5] \\ \\ 18 - 6x + \frac{1}{2}x^2 \text{ for } x \in [5,6] \end{cases} \tag{2.30}$$

For this example, the combination rules of corollary 2 apply so ${}^R\mu_{\tilde{x}^i}(x) = 1 - (1 - F_i(x))^2$ and ${}^L\mu_{\tilde{x}^i}(x) = 1 - (1 - G_i(x))^2$. For example,

$${}^R\mu_{\tilde{x}^1}(x) = \begin{cases} 1 - \left(1 - \frac{1}{2}x^2 - x + \frac{1}{2}\right)^2 & \text{x} \in [1,2] \\ 1 - \left(1 - 3x - \frac{1}{2}x^2 - 3.5\right)^2 & \text{x} \in [2,3] \\ 0 & \text{otherwise} \end{cases} \tag{2.31}$$

is a right possibility distribution function for $\tilde{x}^1$ in the context of $\widetilde{X}$. A left possibility distribution is constructed similarly but using G_i.

2.4 Set Representations

Before examining functions of possibility distributions we consider a method for representing a possibility distribution by a family of nested sets.

These representations, called **set representations**, will be useful when functions of possibility distributions are examined. They are useful for computational as well as theoretical purposes. We will show in the next section that the possibility distribution, of a function of random variables represented by possibility distributions, can be determined by evaluating the function over the set representations of the possibility distributions of the random variables.

Recall that possibility theory can be viewed as a subset of fuzzy set theory. In both theories, membership functions and possibility distribution functions are represented by the symbol μ and in both theories μ maps the universal set X to the unit interval. We begin this section with two definitions from fuzzy set theory that are also used in possibility theory.

Definition 10 (Kaufmann&Gupta [20]) Given a fuzzy set $\tilde{x}$ it's **α-cut** is equal,

$$\tilde{x}_\alpha = \{y \in X \mid \mu_{\tilde{x}}(y) \geq \alpha\} \text{ for } \alpha \in [0,1] \tag{2.32}$$

and it's **strong** $\alpha - cut$ is equal,

$$\tilde{x}_{\alpha+} = \{y \in X \mid \mu_{\tilde{x}}(y) > \alpha\} \text{ for } \alpha \in [0,1] \tag{2.33}$$

Definition 11 $\tilde{}$(Kaufmann&Gupta [20]) Given a fuzzy set $\tilde{x}$, the set $\tilde{x}_{0+} = \{a \mid \mu_{\tilde{x}}(a) > 0\}$ is called the **support** of $\tilde{x}$.

The α-cut for a possibility distribution constructed using the methods of this thesis consists of all elements whose probability of being possible is at least α. The support of such a possibility distribution consists of all elements that have non-zero probability of being possible.

$\alpha-$cuts are used to form a representation for a possibility distribution as a nested family of sets $\{\tilde{x}_\alpha\}_{\alpha\in[0,1]}$. These and similar representations will be used quite extensively in the sequel. This notion can be extended.

Definition 12 (Kruse et al. [23]) The nested family of subsets $\{A_\alpha\}_{\alpha\in[0,1]}$ is called a **set representation** for a possibility distribution $\mu_{\tilde{x}}$ for $\tilde{x}$ if $\forall a \in X$,

$$\mu_{\tilde{x}}(a) = sup\{\alpha \mid a \in A_\alpha\} \tag{2.34}$$

where by definition $\sup\emptyset = 0$.

Set representations have continuity properties.

Theorem 10 (Kruse et al. [23]) Let $(A_\alpha)_{\alpha\in[0,1]}$ be a set representation for a possibility distribution $\mu_{\tilde{x}}$ for x̃. Then $\forall\alpha \in (0,1]$ $\tilde{x}_\alpha = \cap_{\gamma<\alpha} A_\gamma$ and $\forall\alpha \in [0,1)$ $\tilde{x}_{\alpha+}{=}\cup_{\gamma>\alpha} A_\gamma$.

Set representations are bounded above and below by the $\alpha - cuts$ and strong $\alpha - cuts$ of a possibility distribution.

Theorem 11 (Kruse et. al. [23]) The family of sets $(\mathrm{A}_\alpha)_{\alpha\in(0,1)}$ is a set representation for a possibility distribution $\mu_{\tilde{x}}$ for x̃ if and only if it holds that $\tilde{x}_{\alpha+} \subseteq A_\alpha \subseteq \tilde{x}_\alpha$ $\forall\alpha \in (0,1)$.

Example 3 For Example 1, the set representations for ${}^L\mu_{\tilde{x}^1}$ formed by $\alpha-cuts$ and strong $\alpha-cuts$ are as follows, where ${}^L\tilde{x}^1_\alpha$ represents the $\alpha-cut$ of the possibility distribution for $\tilde{x}$ when the possibility distribution for $\tilde{x}$ in the context of $\{\tilde{x}^1, \tilde{x}^2\}$ is ${}^L\mu_{\tilde{x}^1}$:

α	${}^L\tilde{x}^1_\alpha$	α	${}^L\tilde{x}^1_{\alpha+}$
$.925 < \alpha \leq 1$	$\{1\}$	$\alpha = 1$	$\emptyset$
$.4 < \alpha \leq .925$	$\{1, 2\}$	$.925 \leq \alpha < 1$	$\{1\}$
$.25 < \alpha \leq .4$	$\{1, 2, 3\}$	$.4 \leq \alpha < .925$	$\{1, 2\}$
$0 < \alpha \leq .25$	$\{1, 2, 3, 4\}$	$.25 \leq \alpha < .4$	$\{1, 2, 3\}$
$\alpha = 0$	$\{1, 2, 3, 4, 5\}$	$0 \leq \alpha < .25$	$\{1, 2, 3, 4\}$

Example 4 Constructing Set Representations For Example 2, we will form the set representations using the $\alpha - cuts$. Rather than using the possibility distribution functions developed earlier, we will create the set representations from first principles. We begin with the set representations for the left distributions.

Let

$$ {}^L\tilde{x}^1_\alpha = [1, 2\beta + 1] \text{ and } {}^L\tilde{x}^2_\alpha = [4, 2\beta + 4] \text{ for } \beta \in [0, 1]. \tag{2.35}$$

Then the subscript, α as a function of β, can be determined. Noting that the two distributions are identical except for a linear shift, observe that

$$prob(\tilde{x}_i \in^L \tilde{x}^i_\alpha) = \begin{cases} \int_1^{2\beta+1}(x-1)dx = 2\beta^2 \text{ for i=1,2 and } \beta \in [0,.5] \\ \frac{1}{2} + \int_2^{2\beta+1}(3-x)dx = -1+4\beta-2\beta^2 \text{ for } \beta \in [.5,1] \end{cases} . \tag{2.36}$$

Then from (2.13) and (2.24),

$$\alpha = 1 - P(\tilde{x}^1 \in^L \tilde{x}^1_\alpha)P(\tilde{x}^2 \in^L \tilde{x}^2_\alpha) = \begin{cases} 1-(2\beta^2)^2 \text{ for i=1,2 and } \beta \in [0,.5] \\ 1-(-1+4\beta-2\beta^2)^2 \text{ for } \beta \in [.5,1] \end{cases} .$$

Suppose $\alpha = 1-(2\beta^2)^2$, then $\beta = \sqrt{\frac{1}{2}\sqrt{1-\alpha}}$ for $\alpha \in [1,.75]$.

Suppose $\alpha = 1-(-1+4\beta-2\beta^2)^2$, then $\beta = 1 - \frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}$ for $\alpha \in [.75,0]$. Thus a set representation for the left possibility distributions for $\tilde{x}^1$ and $\tilde{x}^2$ in the context of $\{\tilde{x}^1, \tilde{x}^2\}$are given by:

$$^L\tilde{x}^1_\alpha = \begin{cases} [1, 2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)+1] \text{ for } \alpha \in [1,.75] \\ [1, 2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right)+1] \text{ for } \alpha \in [.75,0] \end{cases}$$

and

$$^L\tilde{x}^2_\alpha = \begin{cases} \left[4, 2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)+4\right] \text{ for } \alpha \in [1,.75] \\ \left[4, 2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right)+4\right] \text{ for } \alpha \in [.75,0] \end{cases} .$$

Now let

$$^R\tilde{x}^1_\alpha = [3-2\beta, 3] \text{ and } {}^R\tilde{x}^2_\alpha = [6-2\beta, 6] \text{ for } \beta \in [0,1]. \tag{2.37}$$

Using the same reasoning as above to determine the subscript α as a function of β we get:

$$ {}^{R}\widetilde{x}^{1}_{\alpha} = \begin{cases} [3-2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right),3] \text{ for } \alpha \in [1,.75] \\ [3-2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right),3] \text{ for } \alpha \in [.75,0] \end{cases} $$

and

$$ {}^{R}\widetilde{x}^{2}_{\alpha} = \begin{cases} \left[6-2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right),6\right] \text{ for } \alpha \in [0,.75] \\ \left[6-2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right),6\right] \text{ for } \alpha \in [.75,1] \end{cases} . $$

An alterative construction is to form the $\alpha - cuts$ that concentrate the probability mass. We will use the left superscript U to distinguish this distribution, meaning the points with higher probability density are concentrated in the $\alpha - cuts$ with higher values of α. Let

$$ {}^{U}\widetilde{x}^{1}_{\alpha} = [1+2\beta, 3-2\beta] \text{ and } {}^{U}\widetilde{x}^{2}_{\alpha} = [4+2\beta, 6-2\beta] \text{ for } \beta \in [0,.5]. \quad (2.38) $$

Then the subscript, α as a function of β, can be determined. Noting again that the two distributions are identical except for a linear shift, observe that

$$ prob(\widetilde{x}_i \in^{U} \widetilde{x}^{i}_{\alpha}) = 2\int_{1+2\beta}^{2}(x-1)dx = 1-4\beta^2 \text{ for } \beta \in [0,.5]. \quad (2.39) $$

Then from (2.13) and (2.24),

$$ \alpha = 1 - P(\widetilde{x}^1 \in^{U} \widetilde{x}^{1}_{\alpha})P(X_i \in^{U} \widetilde{x}^{2}_{\alpha}) = 1-(1-4\beta^2)^2 \text{ for } \beta \in [0,.5]. $$

Let $\alpha = 1-(1-4\beta^2)^2$, then $\beta = \frac{1}{2}\sqrt{1-\sqrt{1-\alpha}}$ for $\alpha \in [0,1]$.

Thus set representations for the possibility distributions for $\widetilde{x}^1$ and $\widetilde{x}^2$ in the context of $\{\tilde{x}^1, \tilde{x}^2\}$ are given by:

$$^U\tilde{x}^1_\alpha = \left[1 + \sqrt{1 - \sqrt{1-\alpha}}, 3 - \sqrt{1 - \sqrt{1-\alpha}}\right] \text{ for } \alpha \in [0,1]$$

and

$$^U\tilde{x}^2_\alpha = \left[4 + \sqrt{1 - \sqrt{1-\alpha}}, 6 - \sqrt{1 - \sqrt{1-\alpha}}\right] \text{ for } \alpha \in [0,1].$$

2.5 Functions of Possibilitity Distributions - Extension Principles

Suppose there are two possibility distributions for unknown elements of the real line, $\tilde{x}$ and $\tilde{y}$ and the decision maker is interested in the unknown element $\tilde{x} + \tilde{y}$. In particular, how can the possibility distribution for the sum be determined from the distributions for $\tilde{x}$ and $\tilde{y}$? This section shows how this is done in terms of the contextual possibility distributions. It is also shown how set representations can be used to arrive at the same result. In particular, by applying the operation as a set function on the set representations of each distribution, a set representation for the combined distribution is derived.

The most important property of possibility theory is the ease with which different possibility distributions can be combined. Zadeh provided the formula for constructing the possibility distribution for the function of a possibility distribution as follows:

Definition 13 (Extension Principle) (Zadeh [48]) Let $f : X \to Y$ be an arbitrary function and let $\widetilde{x}$ be a fuzzy set over X. Define a fuzzy set over Y which we denote $f(\widetilde{x})$, by setting $\mu_{f(\widetilde{x})}(b) = sup_{b=f(a)}\mu_{\widetilde{x}}(a)$.

The cumulative probability interpretation of membership values gives justification for this definition. Recall from Theorem 7 that for $B \subset \Psi$,

$$prob(\bigvee_{(\widetilde{x}^i,a)\in B} (\widetilde{x}^i, a) \geq \omega) = \sup\left\{\mu_{\widetilde{x}^i}(a) \mid (\widetilde{x}^i, a) \in B\right\} \tag{2.40}$$

and

$$prob(\bigwedge_{(\widetilde{x}^i,a)\in B} (\widetilde{x}^i, a) \geq \omega) = \inf\left\{\mu_{\widetilde{x}^i}(a) \mid (\widetilde{x}^i, a) \in B\right\}. \tag{2.41}$$

In words, the probability that at least one element of set B is possible is the supremum over the probability that any particular element in B is possible and the probability that all elements of B are possible is the infimum over the probability that any particular element of B is possible.

Definition 14 Given $\left(X, \widetilde{X}, \geq_c\right)$ where $\widetilde{X} = \{\widetilde{x}^i \mid i = 1, \cdots, N\}$ and a measurable function $f : X^N \to Y$. Call y$\in Y$ **possible in the context of** $\widetilde{X}$ if $\exists \prod_{i=}^{N} a_i \in X^N$ such that each a_i is possible in the context of $\widetilde{X}$ and $y = f\left(\prod_{i=}^{N} a_i\right)$.

This definition provides a basis for a probabilistic interpretation of the extension principle.

Theorem 12 Given $\left(X, \widetilde{X}, \geq_c\right)$ where $\widetilde{X} = \{\widetilde{x}^i \mid i = 1, \cdots, N\}$ then a measurable function $f : X^N \to Y$ induces a possibility nest $\geq_p$ on Y for the random vector$f(\prod_{i=1}^N \widetilde{x}^i)$ over Y with contextual possibility distribution function given by the extension principle, i.e. $\forall y \in Y$

$$\mu_{f(\prod_{i=1}^N \widetilde{x}^i)}(y) = \sup_{y=f(\prod_{i=}^N a_i)} \min_{i=1 \text{ to N}} (\mu_{\widetilde{x}^i}(a_i)) \tag{2.42}$$

where $\mu_{f(\prod_{i=1}^N \widetilde{x}^i)}(y) = prob(y$ is possible in the context of $\widetilde{X})$.

Proof:

Let $W = f\left(X^N\right)$. $\forall\, y, w \in W^c$ and $\forall z \in W$ set $y =_p w <_p z$. Define a point to set mapping $\overline{f} : \Psi \to \mathcal{P}\left(\mathcal{Y}\right)$ by

$$\overline{f}((\widetilde{x}^i, a)) = \left\{ \begin{array}{c} y \in Y \mid \exists \prod_{i=}^N a_i \in X^N \text{ with } f\left(\prod_{i=}^N a_i\right) = y \\ \\ \text{and} \\ \\ (\widetilde{x}^i, a) \leq_c \min\left\{(\widetilde{x}^j, a_j) \mid j = 1 \text{ to N}\right\} \end{array} \right\}$$

Since Ψ is linearly ordered, $\left\{\overline{f}((\widetilde{x}^i, a)) \mid (\widetilde{x}^i, a) \in \Psi\right\}$ is nested and thus linearly orders Y. Let $\geq_p$ be this linear ordering. For each $y \in Y$, $f^{-1}(\{w \geq_p y\}) = f^{-1}\left(\overline{f}((\widetilde{x}^i, a))\right)$ for some $(\widetilde{x}^i, a) \in \Psi$. For this $(\widetilde{x}^i, a), \prod_{j=1}^N a_j \in f^{-1}\left(\overline{f}((\widetilde{x}^i, a))\right)$ if and only if $(\widetilde{x}^i, a) \leq_c \min\left\{(\widetilde{x}^j, a_j) \mid j = 1 \text{ to N}\right\}$. Thus $f^{-1}(\{w \geq_p y\}) = \prod_{j=1}^N \{a_j \mid (\widetilde{x}^j, a) \leq_c (\widetilde{x}^i, a_i)\}$ and each set in this product is measurable with respect to the random vector $\widetilde{x}^i$ by assumption on $\geq_c$. Therefore $\{w \geq_p y\}$ is a measurable set in Y.

Let $\widetilde{y} = g(\prod_{i=1}^{N} \widetilde{x}^i)$. Recall that $\omega = \min\{(\widetilde{x}^i, X_i) \mid i = 1, \cdots, N\}$. Then $prob(y \geq_p \widetilde{y}) = prob(\exists \prod_{i=1}^{N} a_i \mid f\left(\prod_{i=1}^{N} a_i\right) = y$ and $\forall\ i = 1, \cdots, N, \omega \leq_c (\widetilde{x}^i, a_i))$. In words, $prob\,(y \geq_p \widetilde{y})$ is the probability that there is an $\prod_{i=1}^{N} a_i$ in X^N such that each a_i is a possible candidate for $\widetilde{x}^i$ in the context of $\widetilde{X}$ and such that $y = g(\prod_{i=1}^{N} \widetilde{x}^i)$, i.e. it is the probability that $g(\prod_{i=1}^{N} \widetilde{x}_i) = y$ is possible. But the probability that each a_i is a possible candidate for $\widetilde{x}^i$ in the context of $\widetilde{X}$ is equal to $\min_{i=1 \text{ to } \mathrm{N}}(\mu_{\widetilde{x}^i}(a_i))$. Also $prob\left(\exists \prod_{i=1}^{N} a_i \mid f\left(\prod_{i=1}^{N} a_i\right) = y\right.$ and $\forall\ i = 1 \ldots N, \omega \leq_c (\widetilde{x}^i, a_i)) = prob(\bigcup_{\prod_{i=1}^{N} a_i \mid f(\prod_{i=1}^{N} a_i) = y} \{\omega \leq_c (\widetilde{x}^i, a_i)\ \forall i = 1, \cdots, N\})$. And these events are nested by the linear ordering of Ψ. Therefore, using property three from the definition of a contextual possibility nest

$$prob\left(\bigcup_{\prod_{i=1}^{N} a_i \mid f(\prod_{i=1}^{N} a_i) = y} \{\omega \leq_c (\widetilde{x}^i, a_i)\ \forall i = 1 \text{ to } N\}\right)$$
$$= \sup\left(\{\omega \leq_c (\widetilde{x}^i, a_i)\ \forall i = 1 \text{ to } N\} \mid f\left(\prod_{i=1}^{N} a_i\right) = y\right)$$
$$= \sup_{y = g(\prod_{i=}^{N} a_i)} \min_{i=1 \text{ to } \mathrm{N}}(\mu_{\widetilde{x}^i}(a_i)). \qquad \square$$

The usefulness of set representations is demonstrated by the following theorem.

Theorem 13 Given $\left(X, \widetilde{X}, \geq_c\right)$ where $\widetilde{X} = \{\widetilde{x}^i \mid i = 1, N\}$. Let $\{A_\alpha^i\}_{\alpha \in [0,1]}$ be a set representation for a possibility distribution for random vector $\widetilde{x}^i$ over X in the context of $\widetilde{X}$ and let $g : X^N \to Y$ be measurable. Then

$\left\{ g(\prod_{i=1}^{N} a_i) \text{ where } a_i \in A_\alpha^i \ \forall\ i = 1, \cdots, N \right\}_{\alpha \in [0,1]}$ is a set representation for the possibilistic variable $g(\prod_{i=1}^{N} \widetilde{x}^i)$ as defined in Theorem 12.

Proof:

Since $\mu_{g(\prod\limits_{i=1}^{N} \widetilde{x}^i)}(y) = \sup_{y=g(\prod\limits_{i=1}^{N} a_i)} \min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i))$, we need to show that $\forall y \in Y$,

$$\sup_{y=g(\prod_{i=1}^{N} a_i)} \min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i))$$
$$= \ sup \left\{ \alpha \mid y \in \left\{ g(\prod_{i=1,N} a_i) \text{ where } a_i \in A_\alpha^i \ \forall i = 1 \cdots N \right\} \right\} \quad (2.43)$$

We consider two cases.

Case one:

Let $y \in Y$. Assume $sup\left\{ \alpha \mid y \in \left\{ g(\prod_{i=1,N} a_i) \text{ where } a_i \in A_\alpha^i \ \forall i = 1, N \right\} \right\} = \tau > 0$. Then $\forall \gamma < \tau$, $y \in \left\{ g(\prod_{i=1,N} a_i) \text{ where } a_i \in A_\gamma^i \ \forall i = 1 \text{ to } N \right\}$ since the A_γ^i's are nested. So $\exists \prod_{i=1}^{N} a_i$ such that $y = g(\prod_{i=1}^{N} a_i)$ and $a_i \in A_\gamma^i$. Thus $\mu_{\widetilde{x}^i}(a_i) \geq \gamma \ \forall i = 1, N$ (definition 2.34) so $\min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i)) \geq \gamma$. But this is true $\forall \gamma < \alpha$ so $\sup_{y=g(\prod_{i=1}^{N} a_i)} \min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i)) \geq \tau$. On the other hand, if $\delta > \tau$ then $y \notin \left\{ g(\prod_{i=1,N} a_i) \text{ where } a_i \in A_\delta^i \ \forall i = 1 \text{ to } N \right\}$. This means that $\forall \prod_{i=1}^{N} a_i$ such that $y = g(\prod_{i=1}^{N} a_i)$, for some j, $a_j \notin A_\delta^j$. Therefore for this j $\mu_{\widetilde{x}^j}(a_j) \leq \delta$ which implies that $\min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i)) \leq \delta$. Since this is true $\forall \delta > \tau$, it must be that $\min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i)) \leq \tau$. Thus $\sup_{y=g(\prod_{i=1}^{N} a_i)} \min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i)) \leq \tau$ and equality follows.

Case two:

Let $y \in Y$. Assume $sup\Big\{\alpha \mid y \in \Big\{g(\prod_{i=1,N} a_i) \text{ where } a_i \in A^i_\alpha \ \forall i = 1, N\Big\}\Big\} = 0$. By the previous argument, $\sup_{y=g(\prod_{i=1}^N a_i)} \min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i)) \leq 0$. But by definition of μ, $\sup_{y=g(\prod_{i=1}^N a_i)} \min_{i=1,N}(\mu_{\widetilde{x}^i}(a_i)) \geq 0.\Box$

This is an important result. It tells us that to evaluate a function of possibility distributions we don't have to evaluate the supremum over the minimums of the membership values. Instead, we can evaluate the function over the $\alpha - cuts$ of the possibility distributions. For some problems, this simplifies to performing interval arithmetic using the $\alpha - cuts$ of each possibility distribution.

Example 5 Consider the random variable $\widetilde{y}=\widetilde{x}^1+\widetilde{x}^2$ where $\widetilde{x}^1$ and $\widetilde{x}^2$ are as given in Example 1. This variable is monotonic in each variable $\widetilde{x}^1$ and $\widetilde{x}^2$. Therefore, we can calculate the set representations formed by the $\alpha-$cuts for the left and right possibility distributions for $\widetilde{y}$ from the $\alpha - cuts$ for the left and right possibility distributions for $\widetilde{x}^1$ and $\widetilde{x}^2$ respectively. The α-cuts for the left possibility distribution are:

α	${}^{L}\tilde{x}^1_\alpha$	${}^{L}\tilde{x}^2_\alpha$	${}^{L}\tilde{y}_\alpha$
$.925 < \alpha \leq 1$	$\{1\}$	$\{1\}$	$\{2\}$
$.775 < \alpha \leq .925$	$\{1,2\}$	$\{1\}$	$\{2,3\}$
$.4 < \alpha \leq .775$	$\{1,2\}$	$\{1,2\}$	$\{2,3,4\}$
$.3 < \alpha \leq .4$	$\{1,2,3\}$	$\{1,2\}$	$\{2,3,4,5\}$
$.125 < \alpha \leq .3$	$\{1,2,3\}$	$\{1,2,3\}$	$\{2,3,4,5,6\}$
$0 < \alpha \leq .125$	$\{1,2,3,4\}$	$\{1,2,3\}$	$\{2,3,4,5,6,7\}$

This has the following meaning;

7.5% of the time $(1-.925)$, $\tilde{y}$=2 is the only possible value for $\tilde{y}$,

22.5% of the time $(1-.775)$, $\tilde{y}$=2 or 3 are the only possible values for $\tilde{y}$,

60% of the time $(1-.4)$, $\tilde{y}$=2,3, or 4 are the only possible values for $\tilde{y}$,

70% of the time $(1-.3)$, $\tilde{y}$=2,3,4 or 5 are the only possible values of $\tilde{y}$,

87.5% of the time $(1-.125)$, $\tilde{y}=2,3,4,5$ or 6 are the only possible values of $\tilde{y}$,

and 100% of the time $\tilde{y}$=2,3,4,5,6 or 7 are the only values for $\tilde{y}$.

The α-cuts for the right possibility distribution are:

α	${}^{R}\widetilde{x}^1_\alpha$	${}^{R}\widetilde{x}^2_\alpha$	${}^{R}\widetilde{y}_\alpha$
$.975 < \alpha \leq 1$	$\{4\}$	$\{3\}$	$\{7\}$
$.95 < \alpha \leq .975$	$\{4, 3\}$	$\{3\}$	$\{6, 7\}$
$.825 < \alpha \leq .95$	$\{4, 3\}$	$\{3, 2\}$	$\{5, 6, 7\}$
$.475 < \alpha \leq .825$	$\{4, 3, 2\}$	$\{3, 2\}$	$\{4, 5, 6, 7\}$
$.25 < \alpha \leq .475$	$\{4, 3, 2\}$	$\{3, 2, 1\}$	$\{3, 4, 5, 6, 7\}$
$0 < \alpha \leq .25$	$\{4, 3, 2, 1\}$	$\{3, 2, 1\}$	$\{2, 3, 4, 5, 6, 7\}$

Example 6 For the set representations from Example 5, consider $\widetilde{y} = (\widetilde{x}^1)^2 - \widetilde{x}^2$. To construct the right possibility distribution for $\widetilde{y}$, we note that over the support of $\widetilde{x}^1$, $\widetilde{y}$ increases as $\widetilde{x}^1$ increases and over the support of $\widetilde{x}^2$, $\widetilde{y}$ is increases as $\widetilde{x}^2$ decreases. Therefore, to construct a set representation for the right distribution for $\widetilde{y}$, we apply interval arithmetic to the set representation for the right distribution for $\widetilde{x}^1$ and the left distribution for $\widetilde{x}^2$. This results in:

$$
{}^{R}\widetilde{y}_{\alpha} = \begin{cases} [\left(3-2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)^2 - \left(2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}+4\right), 5] \text{ for } \alpha \in [1,.75] \\ [\left(3-2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)\right)^2 \\ -\left(2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)+4\right), 5] \text{ for } \alpha \in [.75, 0] \end{cases}
$$

and

$$
{}^{L}\widetilde{y}_{\alpha} = \begin{cases} [-5, \left(2\left(\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)+1\right)^2 - \left(6-2\sqrt{\frac{1}{2}\sqrt{1-\alpha}}\right)] \text{ for } \alpha \in [1,.75] \\ [-5, \left(2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\alpha+1)}\right)}\right)+1\right)^2 \\ -\left(6-2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\alpha+1)}}\right)\right)] \text{ for } \alpha \in [.75, 0] \end{cases}
$$

Using the upper possibility distribution a set representation for another possibility distribution for $\widetilde{y}$ is:

$$
{}^{U}\widetilde{y}_{\alpha} = [\left(1+\sqrt{1-\sqrt{1-\alpha}}\right)^2 - \left(6-\sqrt{1-\sqrt{1-\alpha}}\right),
$$

$$
\left(3-\sqrt{1-\sqrt{1-\alpha}}\right)^2 - \left(4+\sqrt{1-\sqrt{1-\alpha}}\right)] \text{ for } \alpha \in [0,1]
$$

A graph of the upper possibility distribution versus the probability density function for $\widetilde{y}$ is shown in Figure 2.1. A graph of the upper possibility distributions for $\widetilde{x}^1, \widetilde{x}^2$ and $\widetilde{y}$ is shown in Figure 2.2.

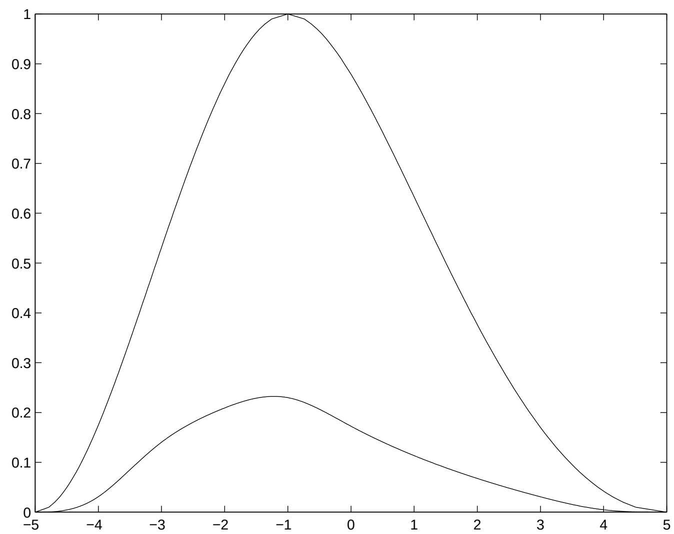

Figure 2.1. Graph of the upper possibility distribution function and probability density function for $\widetilde{y} = (\widetilde{x}^1)^2 - \widetilde{x}^2$.

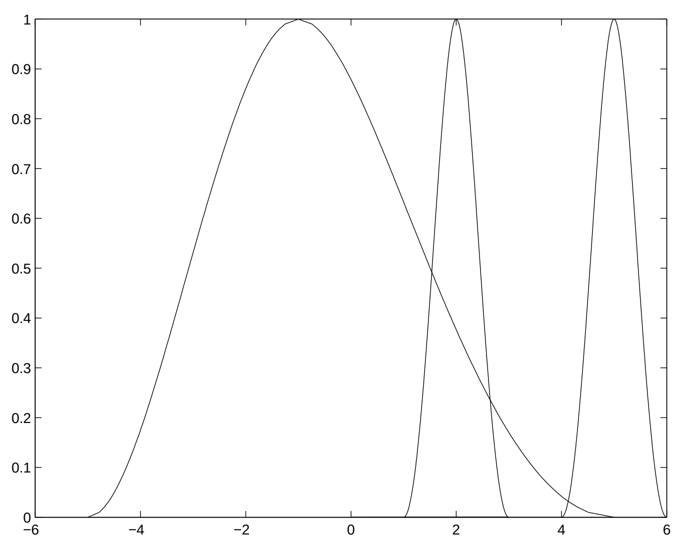

Figure 2.2. Graph of upper possibility distributions for $\widetilde{x}^1$, $\widetilde{x}^2$, and $\widetilde{y} = (\widetilde{x}^1)^2 - \widetilde{x}^2$.

2.6 Probabilistic Based Possibility Distributions for Random Vectors

We consolidate the possibility distributions constructed in the preceding definitions into a single definition.

Definition 15 Let Y be a space and $\widetilde{y}$ a random vector. We will call $\mu_{\widetilde{y}}$ a **probabilistic based possibility distribution** for $\widetilde{y}$ if there exists a contextual possibility nest $\left(X, \widetilde{X}, \geq_c\right)$ and a measurable function $f : X^N \to Y$ such that $\widetilde{y} = f(\prod_{i=1}^{N} \widetilde{x}^i)$ and $\mu_{\widetilde{y}}(a) = prob\left(\widetilde{y} = a \text{ is possible in the context of } \widetilde{X}\right)$.

Note that this definition includes all of the possibility distribution functions constructed earlier. For example, if $\widetilde{X} = \{\widetilde{x}\}$ and $Y = X$ and $f(x) \equiv x$ (i.e. the identity map) then we have the possibility distribution with membership function $\mu_{\widetilde{x}}(a) = prob\,(a \geq_p \widetilde{x})$ (where $\geq_p$ is the restriction of $\geq_c$) and if $\widetilde{X} = \{\widetilde{x}^i \mid i = 1 \text{ to N}\}$ and $f(x) \equiv x$ (i.e. the identity map) then we have the possibility distributions with membership functions $\mu_{\widetilde{x}^i}(a) = prob\,((\widetilde{x}^i, a) \geq_c \omega)$. With this in mind, from this point forward we drop the phrase "in the context of $\widetilde{X}$" from our definition of possibility since it will always be understood to exist.

Using the above definition, we can interpret the measures of possibility theory from a probabilistic point of view.

Theorem 14 Let $\tilde{x}$ be a random vector with probablistic based possibility distribution function $\mu_{\tilde{x}}$. The function $\text{pos}_{\tilde{x}} : \mathcal{P}(X) \to [0,1]$ given by:

$$pos_{\tilde{x}}(A) = prob\left(\bigvee_{a \in A} (a \text{ is possible})\right) \tag{2.44}$$

is the possibility measure associated with the possibility distribution function $\mu_{\tilde{x}}$ (2.24).

Proof:

From Theorem 12 (using the linear ordering of the events (a is possible) and the fact that the probability of the union is the sup of the probabilities from the third property of a possibility nest), $prob\left(\bigvee_{a \in A} (a \text{ is possible})\right) = prob\left(\bigcup_{a \in A} (a \text{ is possible})\right) = \sup_{a \in A} \{prob(a \text{ is possible})\} = \sup_{a \in A} \{\mu_{\tilde{x}}(a)\}$ and the theorem follows from Definition 3. $\square$

In other words, pos(A) is the probability that at least one element of A is possible.

Theorem 15 Let $\tilde{x}$ be a random vector with probablistic based possibility distribution $\mu_{\tilde{x}}$. The function $\text{nec}_{\tilde{x}} : \mathcal{P}(X) \to [0,1]$ given by:

$$nec_{\tilde{x}}(A) = prob\left(\bigwedge_{a \in A^c} (a \text{ is impossible})\right) \tag{2.45}$$

is the necessity measure associated with the possibility distribution function $\mu_{\tilde{x}}$.

Proof:

Again, from Theorem 12,

$$
\begin{aligned}
& prob\left(\bigwedge_{a\in A^c}(a \text{ is impossible})\right) \\
&= 1 - prob\left(\bigvee_{a\in A^c}(a \text{ is possible})\right) \\
&= 1 - \sup_{a\in A^c}\{\mu_{\widetilde{x}}(a)\} \\
&= 1 - pos_{\widetilde{x}}(A^c).
\end{aligned}
$$

Then (2.7) applies.□

In other words, nec$_{\widetilde{x}}$(A) is the probability that every element that is in A^c is impossible, i.e. $\widetilde{x}$ must be in A.

Two other set functions found in the literature on possibility theory also have probabilistic interpretations.

The function $\triangle_{\widetilde{x}} : \mathcal{P}(\mathcal{X}) \rightarrow [0,1]$ is called an **uncertainty measure** and is defined as (see Kruse et.al. [23]),

$$\triangle_{\widetilde{x}}(A) = \inf\{\mu_{\widetilde{x}}(a) \mid a \in A\} = prob\left(\bigwedge_{a\in A}(a \text{ is possible})\right).$$

That is, every element of A has at least probability $\triangle_{\widetilde{x}}(A)$ of being a possible candidate for $\widetilde{x}$.

The function $\bigtriangledown_{\widetilde{x}} : \mathcal{P}(\mathcal{X}) \to [0,1]$ is called a **guaranteed possibility measure** and is defined as (see Kruse et.al. [23]),

$$\bigtriangledown_{\widetilde{x}}(A) = 1 - \triangle_{\widetilde{x}}(A^c) = prob\left(\bigvee_{a \in A^c} (a \text{ is impossible})\right). \qquad (2.46)$$

Thus, every element that is not in A has probability less than $\bigtriangledown_{\widetilde{x}}$ of being a possible candidate for $\widetilde{x}$, or every element with probability at least $\bigtriangledown_{\widetilde{x}}(A)$ of being a possible candidate for $\widetilde{x}$ is in A.

2.7 The Information Contained in a Probabilistic Based Possibility Distribution

Given a probablistic based possibility distribution for a random variable, what can be deduced about the probability distribution of the variable? How can the possibility distribution be used to the advantage of a decision maker? This section examines these questions for the case were $\widetilde{X}$ consists of a finite collection of random variables and where we are interested in the random variable consisting of a function of these random variables.

Let $F_{\widetilde{y}}$ be the cumulative distribution function (c.d.f) for random variable $\widetilde{x}$. We will denote the **support** of $F_{\widetilde{y}}$ by $\text{supp}(F_{\widetilde{y}})$, i.e.

$$\text{supp}(F_{\widetilde{y}}) = \left\{ x \in R \mid dF_{\widetilde{y}}(x) > 0 \right\}.$$

We will use $E(\widetilde{x})$ to denote the expected value of $\widetilde{x}$, i.e. $E(\widetilde{x}) = \int_{-\infty}^{\infty} x dF_{\widetilde{y}}(x)$ if it exists.

Theorem 16 Let $\widetilde{y}$ be a random variable with probabilistic based possibility distribution function $\mu_{\widetilde{y}}$. Let $F_{\widetilde{y}}$ be the c.d.f. for $\widetilde{y}$. Then $\exists\{F_\alpha|\alpha \in (0,1)\}$ where F_α

(1) F_α is a c.d.f.

(2) $\text{supp}(F_\alpha) \subseteq \widetilde{y}_\alpha$

(3) if $\widetilde{y}_\alpha = \widetilde{y}_\beta$ then $F_\alpha = F_\beta$

(4) $F_{\widetilde{y}(y)} = \int_0^1 F_\alpha(y)\, d\alpha$ and

(5) for all A measurable $\int_A \, dF_{\widetilde{y}} = \int_0^1 \int_A \, dF_\alpha \, d\alpha$.

Proof:

Associated with $\widetilde{y}$ is the contextual possibility nest $\left(X, \widetilde{X}, \geq_c\right)$. Let F_ω be the cumulative distribution function for ω as a random vector on Ψ (see (2.12)).

(1)

Let $F_\alpha(y) = F_{\widetilde{y}}(y\,|\,\omega = \gamma(\alpha))$ where $\gamma(\alpha) =_c \inf\{(\widetilde{x}^i, x) \in \Psi \mid F_\omega(\widetilde{x}^i, x) > \alpha\}$ and inf is with respect to the ordering $\geq_c$ (note, if F_ω is continuous then $\gamma(\alpha) = F_\omega^{-1}(\alpha)$). This is well defined since F_ω is right continuous implies that the inf is achieved for some $\gamma(\alpha) \in \Psi$. We also know that the set is non-empty since $F_\omega \to 1$. We also know that this c.d.f. exists from Proposition 4.32 of [2].

(2) Recall that

$$\begin{aligned}
\widetilde{y}_\alpha &= \left\{y \in R \mid \mu_{\widetilde{y}}(y) \geq \alpha\right\} = \{y \in R \mid prob(y \text{ is possible}) \geq \alpha\} \\
&= \left\{y \mid \exists \prod_{i=}^{N} a_i \in X^N \ni y = f\left(\prod_{i=}^{N} a_i\right) \text{ and } \forall i \ prob((\widetilde{x}^i, a_i) \geq_c \omega) \geq \alpha\right\} \\
&= \left\{y \mid \exists \prod_{i=}^{N} a_i \in X^N \ni y = f\left(\prod_{i=}^{N} a_i\right) \text{ and } \forall i \ F_\omega((\widetilde{x}^i, a_i)) \geq \alpha\right\}.
\end{aligned}$$

If $z \notin \widetilde{y}_\alpha$ then $\nexists \prod_{i=}^{N} a_i \in X^N$ such that $z = f\left(\prod_{i=}^{N} a_i\right)$ and $\forall i \ F_\omega((\widetilde{x}^i, a_i)) \geq \alpha$.

But then

$$\begin{aligned}
z &\notin \left\{\begin{array}{c} y \mid \exists \prod_{i=}^{N} a_i \in X^N \text{ such that } y = f\left(\prod_{i=}^{N} a_i\right) \text{ and} \\ \forall i \ (\widetilde{x}^i, a_i) \geq_c \gamma(\alpha) \text{ and for some } j \ (\widetilde{x}^j, a_j) =_c \gamma(\alpha) \end{array}\right\} \\
&= \{y \mid \omega = \gamma(\alpha)\} \text{ where } \omega = \min\left\{(\widetilde{x}^i, X_i) \mid i = 1 \text{ to } N\right\}
\end{aligned}$$

so if $y \notin \widetilde{y}_\alpha$ then $y \neq \widetilde{y}$ and $dF_{\widetilde{y}}(y \mid \omega = \gamma(\alpha)) = 0$.

(3) If $\widetilde{y}_\alpha = \widetilde{y}_\beta$ then $\gamma(\alpha) = \gamma(\beta)$ so $F_\alpha = F_\beta$.

(4) Note that if α is uniformly distributed over (0,1), then the random element $\gamma(\alpha)$ of Ψ has the distribution of ω. To see this consider $F_\omega(\gamma(\alpha)) = prob(\omega \leq_c \gamma(\alpha))$. But

$$(\omega \leq_c \gamma(\alpha)) = \bigcap_{\left(\widetilde{x}^i, x\right) \mid F_\omega\left(\widetilde{x}^i, x\right) > \alpha} \left(\omega \leq_c \left(\widetilde{x}^i, x\right)\right)$$

so by applying property three of a possibility nest we have

$$prob(\omega \leq_c \gamma(\alpha)) = \inf_{\left(\widetilde{x}^i, x\right) \mid F_\omega\left(\widetilde{x}^i, x\right) > \alpha} prob\left(\omega \leq_c \left(\widetilde{x}^i, x\right)\right)$$

$$= \inf_{\left(\widetilde{x}^i,x\right)|F_\omega\left(\widetilde{x}^i,x\right)>\alpha} F_\omega\left(\widetilde{x}^i,x\right)$$
$$= \inf_{F_\omega(\gamma)>\alpha} F_\omega(\gamma).$$

Let $\beta = \inf_{F_\omega(\gamma)>\alpha} F_\omega(\gamma) = F_\omega(\gamma(\alpha))$. Then $prob(\alpha \in [0,\beta]) = F_\omega(\gamma(\alpha))$. Therefore, conditioning on ω,

$$F_{\widetilde{y}}(y) = \int_\Psi F_{\widetilde{y}}(y \mid \omega)dF_\omega = \int_0^1 F_{\widetilde{y}}(y \mid \gamma(\alpha))d\alpha = \int_0^1 F_\alpha(y)d\alpha.$$

(5) For any set A measurable, conditioning on ω, $\int_A dF_{\widetilde{y}} = \int_\Psi \int_A dF_{\widetilde{y}|\omega} dF_\omega = \int_0^1 \int_A dF_\alpha d\alpha$.□

Example 7 Consider the left possibility distribution constructed for $\widetilde{y} = \widetilde{x}^1 + \widetilde{x}^2$ in Example 5. The conditional probabilities are as follows:

Table of $\mathrm{F}_\alpha(r)$						
α	$r = 2$	3	4	5	6	7
$.925 < \alpha \leq 1$	1	1	1	1	1	1
$.775 < \alpha \leq .925$	0	1	1	1	1	1
$.4 < \alpha \leq .775$	0	.3333	1	1	1	1
$.3 < \alpha \leq .4$	0	0	.3750	1	1	1
$.125 < \alpha \leq .3$	0	0	.2857	.8571	1	1
$0 < \alpha \leq .125$	0	0	0	.3000	.8000	1
$\int_0^1 F_\alpha(r)d\alpha$	.0750	.3500	.6875	.8875	.9750	1

This representation of the cumulative distribution function for $\widetilde{y}$ as a parametrized family of conditional probability distributions allows us to estimate various functions of the underlying probability distribution. The following theorem shows that the measures pos and nec provide upper and lower bounds for the probability measure.

Theorem 17 Let pos and nec be the possibility measure and it's associated necessity measure for a probabilistic based possibility distribution for random variable $\widetilde{x}$. Let $A \subseteq X$ be a measurable set. Then

$$nec(A) \leq prob(A) \leq pos(A). \tag{2.47}$$

Proof:

Consider the set $\{\alpha \in (0,1] \mid A \bigcap \widetilde{x}_\alpha \neq \emptyset\}$ where $\widetilde{x}_\alpha$ is the $\alpha - cut$ with respect to the possibility distribution. The case when this set is empty is trivial since, then $nec(A) = prob(A) = pos(A) = 0$. Assume the set is not empty.

We first prove that $prob(A) \leq pos(A)$. Let $\gamma = \sup\{\alpha \in (0,1] \mid A \bigcap \widetilde{x}_\alpha \neq \emptyset\}$. From Theorem 16 $prob(A) = \int_A dF_{\widetilde{y}} = \int_0^1 \int_A dF_\alpha d\alpha$ and $\operatorname{supp}(F_\alpha) \subseteq \widetilde{x}_\alpha$. So if $A \bigcap \widetilde{x}_\alpha = \emptyset$ then $\int_A dF_\alpha = 0$. Then $prob(A) = \int_0^\gamma \int_A dF_\alpha d\alpha$. This integral is maximized if $\forall \alpha < \gamma$, $\operatorname{supp}(F_\alpha) \subseteq A$ in which case $\int_A dF_\alpha = 1$ and $prob(A) = \int_0^\gamma d\alpha = \gamma$. Therefore $prob(A) \leq \gamma$. But

$$
\begin{aligned}
pos(A) &= \sup\{\mu(a) \mid a \in A\} \\
&= \sup\{\sup\{\alpha \mid \alpha \in \widetilde{x}_\alpha\} \mid a \in A\} \\
&= \sup\left\{\alpha \mid A \bigcap \widetilde{x}_\alpha \neq \emptyset\right\} \\
&= \gamma
\end{aligned}
$$

We now show that $nec(A) \leq prob(A)$. Let $\gamma = 1 - \sup\{\alpha \mid A^c \bigcap \widetilde{x}_\alpha \neq \emptyset\}$. Using similar arguments as above, the probability of A is minimized if $\forall\alpha$ such that $A^c \bigcap \widetilde{x}_\alpha \neq \emptyset$ then $\text{supp}(F_\alpha) \bigcap A = \emptyset$ where A^c is the complement of A. Then $\int_A dF(x) = \int_0^1 \int_A dF_\alpha(x) d_\alpha = \int_{1-\gamma}^1 d\alpha = 1-(1-\gamma) = \gamma$. Therefore $\gamma \leq prob(A)$. But $nec(A) = 1 - pos(A^c) = 1 - \sup\{\alpha \mid A^c \bigcap \widetilde{x}_\alpha \neq \emptyset\} = \gamma$.$\square$

This bound can be tightened.

Corollary 3 Let $\widetilde{x}$ be a random variable for which $\exists$ a probabilistic based possibility distribution. Let $\left(X, \widetilde{X}, \geq_c\right)$ be the associated possibility nest. If ? is the set of all probabilistic based possibility distributions corresponding to all orderings $\geq_{c'}$ of $\widetilde{X} \times X$ resulting in a possibility nest then

$$
\sup_{\mu\in\Gamma} nec_\mu(A) \leq prob(A) \leq \inf_{\mu\in\Gamma} pos_\mu(A). \tag{2.48}
$$

where pos_μ and nec_μ are the possibility measure and its associated necessity measure for the possibility distribution μ.

We can use a probabilistic based possibility distribution to calculate a bound on the expected value of the random variable as follows.

Theorem 18 Let $\widetilde{x}$ be a random variable with probabilistic based possibility distribution $\mu_{\widetilde{x}}$ with α-cut $\widetilde{x}_\alpha$ and c.d.f. $F_{\widetilde{x}}$ for which $\text{supp}(F_{\widetilde{x}})$ is bounded. Then

$$E(\widetilde{x}) = \int_0^1 E_\alpha(\widetilde{x}) d\alpha \tag{2.49}$$

and

$$\int_0^1 \inf \widetilde{x}_\alpha d\alpha \leq E(\widetilde{x}) \leq \int_0^1 \sup \widetilde{x}_\alpha d\alpha \tag{2.50}$$

where $E_\alpha(\widetilde{x})$ is the expected value of a random variable $\widetilde{x}$ with c.d.f. F_α of Theorem 15.

Proof:

For the first result, by conditioning on ω as a random element of Ψ and using the results of Theorem 16, we have $E(\widetilde{x}) = E(E(\widetilde{x} \mid \omega)) = \int_\Psi \left(\int_{-\infty}^{\infty} \widetilde{x} dF_{\widetilde{y}|\omega} \right) dF_\omega = \int_0^1 \left(\int_{-\infty}^{\infty} \widetilde{x} dF_\alpha \right) d\alpha$.

For the second result, note first that the integrals are well defined because $\inf \widetilde{x}_\alpha$ and $\sup \widetilde{x}_\alpha$ are bounded ($\text{supp}(F_{\widetilde{x}})$ is assumed bounded) and each is monotonic as functions of α. Then since $\text{supp}(F_\alpha) \subseteq \widetilde{x}_\alpha$, it must hold that $\forall \alpha \inf \widetilde{x}_\alpha \leq E_\alpha(\widetilde{x}) \leq sup \widetilde{x}_\alpha$.$\square$

The following theorem is useful for estimating the variance of the distribution. If $\widetilde{x}$ is a random variable let Var($\widetilde{x}$) be the variance of $\widetilde{x}$.

Theorem 19 Let $\widetilde{x}$ be a random variable with c.d.f. $F_{\widetilde{x}}$ for which $supp\,(F_{\widetilde{x}})$ is bounded. Assume $\exists$ a probabilistic based possibility distribution for $\widetilde{x}$. Then

$$Var(\widetilde{x}) = \int_0^1 E_\alpha(\widetilde{x}^2)d\alpha - \left(\int_0^1 E_\alpha(\widetilde{x})d\alpha\right)^2 \tag{2.51}$$

where $E_\alpha(\widetilde{x})$ is the expected value of a random variable $\widetilde{x}$ with c.d.f. F_α of Theorem 16.

Proof:

Conditioning on ω as a random element of Ψ and using the results of Theorem 16 we have

$$E(\widetilde{x}^2) = E(E(\widetilde{x}^2 \mid \omega)) = \textstyle\int_\Psi \left(\int_{-\infty}^{\infty} \widetilde{x}^2 dF_{\widetilde{y}|\omega}\right) dF_\omega$$
$$= \textstyle\int_0^1 \left(\int_{-\infty}^{\infty} \widetilde{x}^2 dF_\alpha\right) d\alpha = \int_0^1 E_\alpha(x^2)d\alpha.$$

Then since $Var(\widetilde{x}) = E\left(\widetilde{x}^2\right) - E\left(\widetilde{x}\right)^2 = \int_0^1 E_\alpha(\widetilde{x}^2)d\alpha - \left(\int_0^1 E_\alpha(\widetilde{x})d\alpha\right)^2$ the theorem follows. $\square$

For any given α, we may not wish to calculate or may not be able to calculate the probability distribution F_α. Instead, an approximate distribution can be used. A simple approximation for F_α is to assume it is uniformly distributed over $\widetilde{x}_\alpha$. For example, assume that $\widetilde{x}_\alpha = [\widetilde{x}_\alpha^-, \widetilde{x}_\alpha^+]$, a closed interval on the real line. Then the expected value of $\widetilde{x}$ can be approximated as

$$E(\widetilde{x}) = \int_0^1 E_\alpha(\widetilde{x})d\alpha \approx \frac{1}{2}\int_0^1 (\widetilde{x}_\alpha^- + \widetilde{x}_\alpha^+)d\alpha \tag{2.52}$$

and the variance can be approximated as

$$Var(\tilde{x}) \approx \frac{1}{3}\int_0^1((\tilde{x}_\alpha^-)^2 + \tilde{x}_\alpha^+\tilde{x}_\alpha^- + (\tilde{x}_\alpha^+)^2)d\alpha - \frac{1}{4}\left(\int_0^1(\tilde{x}_\alpha^- + \tilde{x}_\alpha^+)d\alpha\right)^2 \tag{2.53}$$

where we used the fact that $\tilde{x}$ uniformly distributed over $[\tilde{x}_\alpha^-, \tilde{x}_\alpha^+]$ implies that $E(\tilde{x}^2) = \frac{1}{3}((\tilde{x}_\alpha^-)^2 + \tilde{x}_\alpha^+\tilde{x}_\alpha^- + (\tilde{x}_\alpha^+)^2)$ and $E(\tilde{x}) = \frac{1}{2}(\tilde{x}_\alpha^- + \tilde{x}_\alpha^+)$ and $Var(\tilde{x}) = E(\tilde{x}^2) - (E(\tilde{x}))^2$. We will not examine the estimates of the variance using possibility distributions further in this thesis. This is an area for future research. From (2.50), we see that this estimate of the expected value is simply the midpoint of the upper and lower bound on the expected value that can be determined from the possibility distribution.

Given a single possibility distribution for a random variable, the estimate of the expected value in (2.52) suggests the following functional.

Definition 16 Let $\tilde{x}$ be a random variable with bounded support. Let $\tilde{x}_\alpha$ be the α-cut for a possibility distribution for $\tilde{x}$. Define the **expected average** of $\tilde{x}$ to be the functional

$$EA(\tilde{x}) = \frac{1}{2}\int_0^1(\tilde{x}_\alpha^- + \tilde{x}_\alpha^+)d\alpha. \tag{2.54}$$

Note that every ordering of ψ that satisfies the property of a possibility nest will produce a different possibility distribution for $\tilde{x}$ and each possibility distribution for $\tilde{x}$ produces a closed interval bounding the expected value of $\widetilde{x}$ (2.50). Therefore, the bound on E($\widetilde{x}$) can be tightened.

Let $\widetilde{y}$ be the random variable $f(\prod_{i=1}^{N} \widetilde{x}_i)$. If ? is the set of all probabilistic based possibility distributions corresponding to all orderings $\geq_c$ of $\widetilde{X} \times X$ resulting in a possibility nest, then

$$E(\widetilde{y}) \in \bigcap_{\mu \in \Gamma} [\int_0^1 \inf\,({}^{\mu}\widetilde{x}_{\alpha})\, d\alpha, \int_0^1 \sup\,({}^{\mu}\widetilde{x}_{\alpha})\, d\alpha]. \tag{2.55}$$

This will suggest another functional estimate of the expected value, but first some additional theory is needed. In the examples presented earlier, we developed left, right possibility distributions. We saw that the possibility measure and its associated necessity measure give upper and lower bounds on the probability measure. In the next theorem, we show that left, right probabilistic based possibility distributions for a random variable give an upper and lower bound on the c.d.f. for the random variable.

Definition 17 Let $\widetilde{x}$ be a real valued random variable with bounded support. A **left possibility distribution** for $\widetilde{x}$ is a possibility distribution ${}^{L}\mu_{\widetilde{x}}$ such that $\forall x, y \in R, x < y$ implies ${}^{L}\mu_{\widetilde{x}}(x) \geq {}^{L}\mu_{\widetilde{x}}(y)$ (i.e. ${}^{L}\mu_{\widetilde{x}}(x)$ is nonincreasing). A **right possibility distribution** for $\widetilde{x}$ is a possibility distribution ${}^{R}\mu_{\widetilde{x}}$ such that $\forall x, y \in R, x < y$ implies ${}^{R}\mu_{\widetilde{x}}(x) \leq {}^{R}\mu_{\widetilde{x}}(y)$ (i.e. ${}^{R}\mu_{\widetilde{x}}(x)$ is nondecreasing).

Theorem 20 For any closed interval $[a, b]$, pos ${}_{{}^{L}\mu_{\tilde{x}}}([a, b]) = {}^{L}\mu_{\tilde{x}}(a)$ and pos ${}_{{}^{R}\mu_{\tilde{x}}}([a, b]) = {}^{R}\mu_{\tilde{x}}(b)$.

Proof: This is clear since (for example) $\text{pos}_{{}^L\mu_{\tilde{x}}}([a,b]) = sup\{{}^L\mu_{\tilde{x}}(x)|$ $x \in [a,b]\}$ and ${}^L\mu_{\tilde{x}}$ nonincreasing implies ${}^L\mu_{\tilde{x}}(a)$ is an upper bound on this set. $\square$

Theorem 21 Let ${}^L\mu_{\widetilde{x}}$ and ${}^R\mu_{\widetilde{x}}$ be probabilistic based possibility distributions for random variable $\widetilde{x}$ with c.d.f. $F_{\widetilde{x}}$, where supp$(F_{\widetilde{x}})$ is bounded. Then

$$1 -^L \mu_{\widetilde{x}}(x) \leq F_{\widetilde{x}}(x) \leq^R \mu_{\widetilde{x}}(x). \tag{2.56}$$

Proof:

Note that ${}^L\mu_{\widetilde{x}}(x)$ gives the possibility that $\widetilde{x} \in [x, \widetilde{x}_{0+}^{+}]$ so by Theorem 17 $prob\left(\widetilde{x} \in (x, \widetilde{x}_{0+}^{+}]\right) \leq prob\left(\widetilde{x} \in [x, \widetilde{x}_{0+}^{+}]\right) \leq pos\left([x, \widetilde{x}_{0+}^{+}]\right) =^L \mu_{\widetilde{x}}(x)$. But $prob\left(\widetilde{x} \in (x, \widetilde{x}_{0+}^{+}]\right) = 1 - prob\left(\widetilde{x} \in [\widetilde{x}_{0+}^{-}, x]\right) = 1 - F_{\widetilde{x}}(x)$ so $F_{\widetilde{x}}(x) \geq 1 -^L \mu_{\widetilde{x}}(x)$. On the other hand, ${}^R\mu_{\widetilde{x}}$ gives the possibility that $\widetilde{x} \in [\widetilde{x}_{0+}^{-}, x]$ which by Theorem 17 implies that $F_{\widetilde{x}}(x) = prob\left(\widetilde{x} \in [\widetilde{x}_{0+}^{-}, x]\right) \leq^R \mu_{\widetilde{x}}(x)$.$\square$

We can use this result to estimate the distribution function of $\widetilde{x}$ and the expected value of $\widetilde{x}$.

Theorem 22 Let ${}^L\mu_{\widetilde{x}}(x)$ and ${}^R\mu_{\widetilde{x}}(x)$ be probabilistic based possibility distributions for random variable $\widetilde{x}$ with bounded support. If FE(x) is given by the formula

$$FE(x) = \lim_{\epsilon \to 0, \epsilon > 0} \frac{1}{2}\left(1 -^L \mu_{\widetilde{x}}(x+\epsilon) +^R \mu_{\widetilde{x}}(x+\epsilon)\right) \tag{2.57}$$

then FE(x) is a cumulative distribution function and the expected value of the random variable X represented by FE(x) is given by the formula

$$E(X) = \frac{1}{2}\int_0^1 \left({}^L\widetilde{x}_\alpha^+ +^R \widetilde{x}_\alpha^-\right) d\alpha. \tag{2.58}$$

Proof: Note that ${}^R\mu_{\widetilde{x}}(x)$ and ${}^L\mu_{\widetilde{x}}(x)$ are monotone so the right-hand limit exists and FE(x) is well-defined.

Let $x < y$, then ${}^R\mu_{\widetilde{x}}(x) \leq^R \mu_{\widetilde{x}}(y)$ and ${}^L\mu_{\widetilde{x}}(y) \leq^R \mu_{\widetilde{x}}(x)$ so that $1 -^L \mu_{\widetilde{x}}(x) \leq 1 -^L \mu_{\widetilde{x}}(y)$. Then $\frac{1}{2}\left(1 -^L \mu_{\widetilde{x}}(x) +^R \mu_{\widetilde{x}}(x)\right) \leq \frac{1}{2}\left(1 -^L \mu_{\widetilde{x}}(y) +^R \mu_{\widetilde{x}}(y)\right)$ thus FE is monotone increasing. Also, ${}^L\mu_{\widetilde{x}}(y) \to 0$ as x$\to \infty$ and ${}^L\mu_{\widetilde{x}}(y) \to 1$ as x$\to -\infty$, and ${}^R\mu_{\widetilde{x}}(y) \to 1$ as x$\to \infty$ and ${}^R\mu_{\widetilde{x}}(y) \to 0$ as x$\to -\infty$. This gives FE(x)$\to 1$ as x$\to \infty$ and FE(x)$\to 0$ as x$\to -\infty$. FE is continuous from the right by definition. Therefore, FE is a distribution function. Let X be the random variable represented by FE. Then $E(X) = \widetilde{x}_{0+}^- + \int_{\widetilde{x}_{0+}^-}^{\widetilde{x}_{0+}^+} (1 - FE(x))\, dx = \widetilde{x}_{0+}^- + \int_{\widetilde{x}_{0+}^-}^{\widetilde{x}_{0+}^+} \left(1 - \frac{1}{2}\left(1 -^L \mu_{\widetilde{x}}(x) +^R \mu_{\widetilde{x}}(x)\right)\right) dx$ which reduces to

$$\begin{gathered}
\widetilde{x}_{0+}^- + \int_{\widetilde{x}_{0+}^-}^{\widetilde{x}_{0+}^+} \left(\tfrac{1}{2} + \tfrac{1}{2}{}^L\mu_{\widetilde{x}}(x) - \tfrac{1}{2}{}^R\mu_{\widetilde{x}}(x)\right) dx \\
= \widetilde{x}_{0+}^- + \tfrac{1}{2}\left(\widetilde{x}_{0+}^+ - \widetilde{x}_{0+}^-\right) + \tfrac{1}{2}\int_{\widetilde{x}_{0+}^-}^{\widetilde{x}_{0+}^+} \left({}^L\mu_{\widetilde{x}}(x)\right) dx - \tfrac{1}{2}\int_{\widetilde{x}_{0+}^-}^{\widetilde{x}_{0+}^+} \left({}^R\mu_{\widetilde{x}}(x)\right) dx \\
\text{but } \int_{\widetilde{x}_{0+}^-}^{\widetilde{x}_{0+}^+} \left({}^L\mu_{\widetilde{x}}(x)\right) dx = \int_0^1 \left({}^L\widetilde{x}_\alpha^+\right) d\alpha - \widetilde{x}_0^- \\
\text{and } \int_{\widetilde{x}_{0+}^-}^{\widetilde{x}_{0+}^+} \left({}^R\mu_{\widetilde{x}}(x)\right) dx = \widetilde{x}_{0+}^+ - \int_0^1 \left({}^R\widetilde{x}_\alpha^-\right) d\alpha.
\end{gathered}$$

Substitution gives the desired result. Note that these last two relationships follow from consideration of the area of the rectangle with base $[\widetilde{x}_{0+}^-, \widetilde{x}_{0+}^+]$ and height one and the areas under the curves ${}^L\mu_{\tilde{x}}(x)$ and ${}^R\mu_{\tilde{x}}(x)$ in $R \times [0,1]$ and the curves ${}^L\tilde{x}_\alpha^+$ and ${}^R\tilde{x}_\alpha^-$ in $[0,1] \times R$.□

This gives us an alternative estimate of the expected value for those situations where left, right probabilistic based possibility distributions are available or computable.

Definition 18 Let ${}^L\mu_{\tilde{x}}(x)$ and ${}^R\mu_{\tilde{x}}(x)$ be possibility distributions for random variable $\tilde{x}$ with bounded support. Define the **estimated expectation** to be the functional

$$EE(\tilde{x}) = \frac{1}{2}\int_0^1 \left({}^L\tilde{x}_\alpha^+ +{}^R\tilde{x}_\alpha^-\right) d\alpha. \tag{2.59}$$

It turns out that in some special situations this estimate of the expected value is exact. Defining these circumstances is an area of future research.

Example 8 For Example 1, the probability distributions for $\tilde{x}^1$, $\tilde{x}^2$ and $\tilde{y} =$

$\tilde{x}^1 + \tilde{x}^2$ are

r	$P(\tilde{x}^1 = r)$	$P(\tilde{x}^2 = r)$	$P(\tilde{y}$=r)
1	.25	.3	0
2	.5	.5	$(.25)(.3) = .075$
3	.125	.2	$(.25)(.5) + (.5)(.3) = .275$
4	.125	0	$(.5)(.5) + (.125)(.3) + (.25)(.2) = .3375$
5	0	0	$(.125)(.5) + (.5)(.2) + (.125)(.3) = .2$
6	0	0	$(.125)(.5) + (.125)(.2) = .0875$
7	0	0	$(.125)(.2) = .025$

The expected value of $\tilde{y}$ is

$$E(\tilde{y}) = 2(.075) + 3(.275) + 4(.3375) + 5(.2) + 6(.0875) + 7(.025) = 4.025.$$

Recall that the left, right distributions for $\tilde{y}$ are:

α	${}^L\tilde{y}_\alpha$
$.925 < \alpha \leq 1$	$\{2\}$
$.775 < \alpha \leq .925$	$\{2,3\}$
$.4 < \alpha \leq .775$	$\{2,3,4\}$
$.3 < \alpha \leq .4$	$\{2,3,4,5\}$
$.125 < \alpha \leq .3$	$\{2,3,4,5,6\}$
$0 < \alpha \leq .125$	$\{2,3,4,5,6,7\}$

α	${}^R\tilde{y}_\alpha$
$.975 < \alpha \leq 1$	$\{7\}$
$.95 < \alpha \leq .975$	$\{6,7\}$
$.825 < \alpha \leq .95$	$\{5,6,7\}$
$.475 < \alpha \leq .825$	$\{4,5,6,7\}$
$.25 < \alpha \leq .475$	$\{3,4,5,6,7\}$
$0 < \alpha \leq .25$	$\{2,3,4,5,6,7\}$

The upper estimate of the expected value, $\int_0^1 \left({}^L\widetilde{x}_\alpha^+\right) d\alpha$, is

$$\begin{aligned} & 2(1 - .925) + 3(.925 - .775) + 4(.775 - .4) \\ + \quad & 5(.4 - .3) + 6(.3 - .125) + 7(.125) = 4.525. \end{aligned}$$

The lower estimate of the expected value, $\int_0^1 \left({}^R\widetilde{x}_\alpha^+\right) d\alpha$, is

$$\begin{aligned} & 7(1 - .975) + 6(.975 - .95) \\ + \quad & 5(.95 - .825) + 4(.825 - .475) + 3(.475 - .25) + 2(.25) = 3.525. \end{aligned}$$

The estimated expected value is:

$$EE(\widetilde{y}) = \frac{1}{2}(4.525 + 3.525) = 4.025.$$

The EE functional does not always equal the actual expected value. For example, consider the random variable $\widetilde{y} = \widetilde{x}^1 * \widetilde{x}^2$.

r	$P(\widetilde{x}^1 = r)$	$P(\widetilde{x}^2 = r)$	$P(\widetilde{y}=r)$
1	.25	.3	$(.25)(.3) = .075$
2	.5	.5	$(.5)(.3) + (.25)(.5) = .275$
3	.125	.2	$(.125)(.3) + (.25)(.2) = .0875$
4	.125	0	$(.5)(.5) + (.125)(.3) = .2875$
6	0	0	$(.125)(.5) + (.5)(.2) = .1625$
8	0	0	$(.125)(.5) = .0625$
9	0	0	$(.125)(.2) = .025$
12	0	0	$(.125)(.2) = .025$

The expected value of $\widetilde{y}$ is

$$\begin{aligned} & 1(.075) + 2(.275) \\ + \quad & 3(.0875) + 4(.2875) + 6(.1625) + 8(.0625) + 9(.025) + 12(.025) = 4.0375. \end{aligned}$$

The left, right distributions for $\widetilde{y}$ and the upper, lower estimates of the expected value are:

α	${}^{L}\widetilde{x}^{1}_{\alpha}$	${}^{L}\widetilde{x}^{2}_{\alpha}$	${}^{L}\widetilde{y}_{\alpha}$
$.925 < \alpha \leq 1$	$\{1\}$	$\{1\}$	$\{1\}$
$.775 < \alpha \leq .925$	$\{1, 2\}$	$\{1\}$	$\{1, 2\}$
$.4 < \alpha \leq .775$	$\{1, 2\}$	$\{1, 2\}$	$\{1, 2, 4\}$
$.3 < \alpha \leq .4$	$\{1, 2, 3\}$	$\{1, 2\}$	$\{1, 2, 3, 4, 6\}$
$.125 < \alpha \leq .3$	$\{1, 2, 3\}$	$\{1, 2, 3\}$	$\{1, 2, 3, 4, 6, 9\}$
$0 < \alpha \leq .125$	$\{1, 2, 3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 3, 4, 6, 8, 9, 12\}$

The upper estimate of the expected value, $\int_0^1 \left({}^{L}\widetilde{x}^{+}_{\alpha}\right) d\alpha$, is

$$
\begin{aligned}
& 1(1 - .925) + 2(.925 - .775) \\
+ \quad & 4(.775 - .4) + 6(.4 - .3) + 9(.3 - .125) + 12(.125) = 5.55.
\end{aligned}
$$

α	${}^{R}\widetilde{x}^{1}_{\alpha}$	${}^{R}\widetilde{x}^{2}_{\alpha}$	${}^{R}\widetilde{y}_{\alpha}$
$.975 < \alpha \leq 1$	$\{4\}$	$\{3\}$	$\{12\}$
$.95 < \alpha \leq .975$	$\{4, 3\}$	$\{3\}$	$\{9, 12\}$
$.825 < \alpha \leq .95$	$\{4, 3\}$	$\{3, 2\}$	$\{6, 8, 9, 12\}$
$.475 < \alpha \leq .825$	$\{4, 3, 2\}$	$\{3, 2\}$	$\{4, 6, 8, 9, 12\}$
$.25 < \alpha \leq .475$	$\{4, 3, 2\}$	$\{3, 2, 1\}$	$\{2, 3, 4, 6, 8, 9, 12\}$
$0 < \alpha \leq .25$	$\{4, 3, 2, 1\}$	$\{3, 2, 1\}$	$\{1, 2, 3, 4, 6, 8, 9, 12\}$

The lower estimate of the expected value, $\int_0^1 \left({}^R\widetilde{x}_\alpha^+\right) d\alpha$, is

$$
\begin{aligned}
& 12(1-.975)+9(.975-.95) \\
+ \quad & 6(.95-.825)+4(.825-.475)+2(.475-.25)+1(.25)=3.375.
\end{aligned}
$$

Then $\mathrm{EE}(\widetilde{y}) = \frac{1}{2}(3.375 + 5.55) = 4.4625$.

Note that in this case we do not get the expected value.

Example 9 The distribution of $\widetilde{y}$ for Example 2 was shown in the introduction to this thesis were we saw that $E(\widetilde{y}) = -.8333 = EE(\widetilde{y})$. Using the upper possibility distribution of Example 6 the EA estimate of the expected value is, after some simplification:

$$
EA(\widetilde{y}) = \frac{1}{2}\int_0^1 \left(2 - 4\sqrt{1-\sqrt{1-\beta}} - 2\sqrt{1-\beta}\right) d\beta = -.73333.
$$

An upper and lower bound on the distribution function of $\widetilde{y}$, derived from the left, right possibility distributions is shown in Figure 2.3.

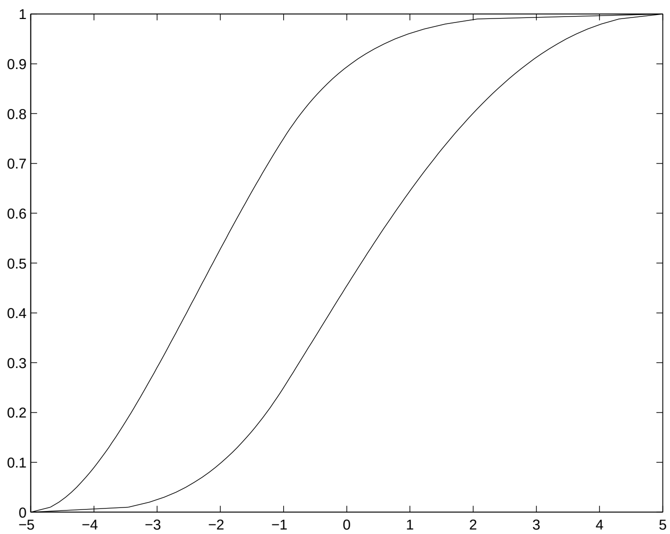

Figure 2.3. Graph of upper and lower bound on cumulative distribution of $\widetilde{y} = (\widetilde{x}^1)^2 - \widetilde{x}^2$.

Example 10 Consider $Z = X^2Y$ where X and Y have the distributions of example 2. Then $E(Z) = E(X^2)E(Y)$ and $E(X^2)=\int_1^2 x^2(x-1)dx + \int_2^3 x^2(3-x)dx = 4.1667$ and $E(Y) = \int_4^5 y(y-4)dy + \int_5^6 y(6-y)dy = 5.0$ so $E(Z) = (4.1667)\,(5) = 20.834$

The EE functional estimate is calculated as follows:

$$Z_\beta^R = \begin{cases} [\left(3-2\sqrt{\frac{1}{2}\sqrt{1-\beta}}\right)^2 \left(6-2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\beta+1)}\right)}\right)\right), 108] \\ \text{for } \beta \in [1, .75] \\ [\left(3-2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\beta+1)}}\right)\right)^2 \left(6-2\left(\sqrt{\frac{1}{2}\sqrt{1-\beta}}\right)\right), 108] \\ \text{for } \beta \in [.75, 0] \end{cases}$$

and

$$Z_\beta^L = \begin{cases} [16, \left(2\left(\sqrt{\frac{1}{2}\sqrt{1-\beta}}\right)+1\right)^2 \left(2\left(\sqrt{\frac{1}{2}\sqrt{1-\beta}}\right)+4\right)] \text{ for } \beta \in [1, .75] \\ [16, \left(2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\beta+1)}\right)}\right)+1\right)^2 * \\ \left(2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\beta+1)}\right)}\right)+4\right)] \text{ for } \beta \in [.75, 0] \end{cases}$$

Then the integral over the left endpoints of the right distribution (lower bound) is

$$\begin{aligned} &\int_{.75}^{1} \left(\left(3-2\sqrt{\frac{1}{2}\sqrt{1-\beta}}\right)^2 \left(6-2\left(\sqrt{\frac{1}{2}\sqrt{1-\beta}}\right)\right)\right) d\beta \\ &+ \int_0^{.75} \left(\left(3-2\left(1-\frac{1}{2}\sqrt{2-2\sqrt{(-\beta+1)}}\right)\right)^2\right. \\ &* \left.\left(6-2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\beta+1)}\right)}\right)\right)\right) d\beta \\ &= 15.8142 \end{aligned}$$

The integral over the right endpoints of the left distribution (upper bound) is

$$\begin{aligned} &\int_{.75}^{1} \left(\left(2\left(\sqrt{\frac{1}{2}\sqrt{1-\beta}}\right)+1\right)^2 \left(2\left(\sqrt{\frac{1}{2}\sqrt{1-\beta}}\right)+4\right)\right) d\beta \\ &+ \int_0^{.75} \left(\left(2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\beta+1)}\right)}\right)+1\right)^2\right. \\ &* \left.\left(2\left(1-\frac{1}{2}\sqrt{\left(2-2\sqrt{(-\beta+1)}\right)}\right)+4\right)\right) d\beta \\ &= 27.1859 \end{aligned}$$

giving $EE(Z) = \frac{1}{2}(15.8142 + 27.1859)\,. = 21.5$

Using the EA functional estimate, the possibility distribution for Z using the upper possibility distributions for X and Y is:

$$\begin{aligned} & Z_\beta^U \\ & = \left[\left(1+\sqrt{1-\sqrt{1-\beta}}\right)^2\left(4+\sqrt{1-\sqrt{1-\beta}}\right),\right. \\ & \left.\left(3-\sqrt{1-\sqrt{1-\beta}}\right)^2\left(6-\sqrt{1-\sqrt{1-\beta}}\right)\right] \\ & \text{for } \beta \in [0,1]. \end{aligned}$$

Then

$$\begin{aligned} & EA(Z) \\ = \;& \frac{1}{2}\int_0^1 \left(\left(1+\sqrt{1-\sqrt{1-\beta}}\right)^2\left(4+\sqrt{1-\sqrt{1-\beta}}\right)\right. \\ + \;& \left.\left(3-\sqrt{1-\sqrt{1-\beta}}\right)^2\left(6-\sqrt{1-\sqrt{1-\beta}}\right)\right) d\beta \\ = \;& 22.4. \end{aligned}$$

Example 11 Let f(x)=$\frac{1}{1+x}$ and $f_{\widetilde{x}}(x) = e^{-x}$ for x$\in$ $[0, 20]$ (we truncate the distribution so the support is bounded). Note that the support of $\widetilde{x}$ is the interval [0,20]. Then E($\widetilde{x}$)=$\int_0^{20} e^{-x}\left(\frac{1}{1+x}\right) dx = .59635$.

Let ${}^R\widetilde{x}_\beta = [20\alpha, 20]$ for $\alpha \in [0,1]$. Then $\text{prob}(X \in [20\alpha, 20]) = \int_{20\alpha}^{20} e^{-x} dx = -e^{-20} + e^{-20\alpha}$ so $\beta = 1 - (-e^{-20} + e^{-20\alpha}) = 1 + e^{-20} - e^{-20\alpha}$ and $\alpha = -\frac{1}{20}\ln\left(-\beta + 1 + e^{-20}\right)$. Therefore ${}^R\widetilde{x}_\beta = [20\left(-\frac{1}{20}\ln\left(-\beta+1+e^{-20}\right)\right), 20]$ or

$${}^R\widetilde{x}_\beta = [-\ln\left(-\beta + 1 + e^{-20}\right), 20] \text{ for } \beta \in [0,1]$$

Now let ${}^{L}\widetilde{x}_{\beta} = [0, 20 - 20\alpha]$. Then prob$(X \in [0, 20 - 20\alpha]) = \int_0^{20-20\alpha} e^{-x} dx = -e^{-20+20\alpha} + 1$ so $\beta = 1 - (-e^{-20+20\alpha} + 1) = e^{-20+20\alpha}$ and $\alpha = \frac{1}{20} \ln \beta + 1$. Therefore ${}^{L}\widetilde{x}_{\beta} = [0, 20 - 20\left(\frac{1}{20} \ln \beta + 1\right)]$ or

$$ {}^{L}\widetilde{x}_{\beta} = [1, -\ln(\beta)] $$

Now for $\widetilde{y}$=f($\widetilde{x}$), note that for the interval [a,b], $\frac{1}{1+[a,b]} = [\frac{1}{1+b}, \frac{1}{1+a}]$ so

$$ {}^{R}\widetilde{y}_{\beta} = \left[\frac{1}{21}, \left(1 - \ln\left(-\beta + 1 + e^{-20}\right)\right)^{-1}\right] \text{ for } \beta \in [0,1] $$

An upper bound of E($\widetilde{y}$) is $\int_0^1 \left((1 - \ln(-\beta + 1 + e^{-20}))^{-1}\right) d\beta = .59635$.

$$ {}^{L}\widetilde{y}_{\beta} = \left[(1 - \ln(\beta))^{-1}, 1\right] $$

A lower bound on E($\widetilde{y}$) is $\int_0^1 \left((1 - \ln(\beta))^{-1}\right) d\beta = .59635$

The upper bound equals the lower bound equals the expected value as it must when there is a single random variable in the context. In this case left and right possibility distributions are just cumulative probability distributions.

In the next chapter we examine single possibility distribution representations of random variables and a norm on these representations motivated by the EA functional. This allows us to examine the convergence properties of estimating expected values using this functional. Future research will focus on the EE functional and what it implies. We anticipate that both functionals will ultimately have their place in applications.

3. The Space of Fuzzy Numbers

In this chapter we examine the space of fuzzy numbers. This space includes possibility distributions for real valued random variables with bounded supports. However, the space of fuzzy numbers is a more general concept.

3.1 Fuzzy Numbers

In this chapter the set of fuzzy sets over the real line is considered (for example random variables represented by single possibility distributions). Consideration will be further limited to fuzzy sets that capture the idea of a fuzzy measurement of a real quantity. We wish to capture the idea of an approximate number such as "a number close to one". It is desirable to have certain properties hold when defining fuzzy quantities and relationships on them. This may be for intuitive as well as practical reasons. In particular, at a given α-level of possibility, convexity offers computational advantage. For certain distributions it also has intuitive appeal. For example, if it is equally likely that a measurement taking a value in R is x or y, then any z between x and y should be of equal or greater possibility. It does not seem too restrictive to expect that if there exists a quantity arbitrarily close to x that is at least

α possible, then x is at least α possible. Finally, the possible values that we believe a single finite object can be should be bounded if we assume there exists a single correct measurement in the underling reality. The following standard definition is given.

Recall that fuzzy set $\widetilde{x}$ is characterized by a membership function $\mu_{\widetilde{x}} : X \to [0,1]$. $\widetilde{x}$ is called normal if $\exists\ x \in X \ni \mu_{\widetilde{x}}(x) = 1$ and $\widetilde{x}_\alpha = \{x \in X | \mu_{\widetilde{x}}(x) \geq \alpha\}$.

Definition 19 (Klir&Yuan [21]) A **fuzzy number**, x̃, is a fuzzy subset of R such that 1) x̃ is normal 2) $\forall \alpha \in (0,1]$ $\tilde{\mathrm{x}}_\alpha$ is a closed interval and 3) $\widetilde{x}_{0+}$ is bounded.

Two special types of fuzzy numbers are often found in the literature as follows.

Definition 20 A **trapezoidal fuzzy number** is a fuzzy number $\widetilde{x}$ characterized by the set (a,b,c,d) where a,b,c,d are real numbers with a$\leq b \leq c \leq d$ and $\widetilde{x}_\alpha = [a + \alpha(b - a), d + \alpha(c - d)]$. A **triangular fuzzy number** is a fuzzy number $\widetilde{x}$ characterized by the triple (a,b,c) where a,b,c are real numbers with a$\leq b \leq c$ and $\widetilde{x}_\alpha = [a + \alpha(b - a), c + \alpha(b - c)]$.

Let $\mathcal{E}$ be the space of fuzzy numbers. The extension principle is used to define mathematics over $\mathcal{E}$ (see Definition 13).

Definition 21 (Fuzzy Arithmetic) Let R be the real line, and $\widetilde{x}$ and $\widetilde{y}$ be two fuzzy numbers over R. Let * equal +,-,$\times$ or $\div$. Define a fuzzy set for the quantity $\widetilde{x} * \widetilde{y}$ by the following formula:

$$\mu_{\widetilde{x}*\widetilde{y}}(c) = \sup_{a*b=c} \left\{ \min \left\{ \mu_{\widetilde{x}}(a), \mu_{\widetilde{y}}(b) \right\} \right\} \tag{3.1}$$

where sup$\{\emptyset\}$=0 and in the case where $\mu_{\widetilde{y}}(0) > 0$ then $\widetilde{x} \div \widetilde{y}$ is left undefined.

Theorem 23 (for proof see Klir&Yuan [21]) $\mathcal{E}$ is closed under the operations of +,-,$\times$ and $\div$ (provided it is defined).

Example 12 Consider the fuzzy numbers defined as follows:

$$\mu_{\widetilde{x}}(a) = \begin{cases} a-2 & a \in [2,3] \\ 4-a & a \in [3,4] \\ 0 & otherwise \end{cases} \qquad \mu_{\widetilde{y}}(b) = \begin{cases} b-4 & b \in [4,5] \\ 6-b & b \in [5,6] \\ 0 & otherwise \end{cases}$$

Then the fuzzy number $\widetilde{z} = \widetilde{x} + \widetilde{y}$ is defined as follows:

$$\mu_{\widetilde{z}}(c) = \begin{cases} .5(c-6) & c \in [6,8] \\ .5(10-c) & c \in [8,10] \\ 0 & otherwise \end{cases}$$

Applying Definition 12 to obtain a set representation for $\widetilde{x}$ we get [2+α,4-α]$_{\alpha\in[0,1]}$ and for $\widetilde{y}$ [4+α,6-α]$_{\alpha\in[0,1]}$. Therefore, a set representation for $\widetilde{x}+\widetilde{y}$ is given by using interval arithmetic to arrive at [6+2α,10-2α]$_{\alpha\in[0,1]}$. However, as mentioned before, fuzzy numbers are more general than interval numbers since each fuzzy number is a weighted family of intervals, (probability weighted when the distribution is a possibility distribution as constucted in chapter two).

3.2 A Normed Space of Fuzzy Number Equivalence Classes

Suppose that all a decision maker knows about some unknown quantity is that it is an element in an interval, [a,b]. The decision maker might indicate the amount he would pay in exchange for the unknown amount. With no additional information, the decision maker might assume a uniform probability distribution over [a,b] and choose the midpoint of the range, i.e. the expected value. Another way the decision maker might approach the problem is to find the number which balances the maximum possible gain against the maximum possible loss. For example, if the decision maker pays 1.5 in exchange for the interval of possible values [1,2] then the largest possible gain of .5 (2-1.5) is equal to the maximum possible loss of .5 (1.5-1). Using this logic, the decision maker's utility for an interval of possible values is the midpoint of the interval and the decision maker would be neutral with respect to two intervals with the same midpoint. The extension of this idea to fuzzy numbers involves $\alpha - cuts$, which is considered next.

Let $\mathcal{E}$={x̃ | x̃ is a fuzzy number}. An equivalence relationship on $\mathcal{E} \times \mathcal{E}$ is defined as follows.

Definition 22 Let $\tilde{x}_\alpha$= $[\tilde{x}_\alpha^-, \tilde{x}_\alpha^+]$ and $\tilde{y}_\alpha$=$[\tilde{y}_\alpha^-, \tilde{y}_\alpha^+]$. Then $\tilde{x} \equiv \tilde{y}$ iff $\forall \alpha \in (0,1]$ $\exists \epsilon_\alpha \in$R such that $[\tilde{x}_\alpha^- - \epsilon_\alpha, \tilde{x}_\alpha^+ + \epsilon_\alpha] = [\tilde{y}_\alpha^-, \tilde{y}_\alpha^+]$.

It is clear that this is an equivalence relation (see Diamond [7]). If a decision maker's utility for an interval is the midpoint of the interval, then this decision maker is neutral with respect to every $\alpha - cut$ of all members of an equivalence class. Let $\mathcal{L}$ be the partition of $\mathcal{E}$ into equivalence classes resulting from this relationship. If $\tilde{x} \in \mathcal{E}$ then $< \tilde{x} >$ will denote the element of $\mathcal{L}$ consisting of all fuzzy numbers equivalent to $\tilde{x}$. In other words, define two fuzzy numbers to be equivalent if the midpoint of each α-cut is the same for the two fuzzy sets.

Theorem 24 Let $< \tilde{x} >\in \mathcal{L}$, then $\exists \widetilde{m} \in< \tilde{x} >$ **of minimal possibility**, i.e. if $\widetilde{m}_\alpha$=$[\widetilde{m}_\alpha^-,\widetilde{m}_\alpha^+]$ and $\tilde{y}\in< \tilde{x} >$ with $\tilde{y}_\alpha = [\tilde{y}_\alpha^-,\tilde{y}_\alpha^+]$ then $\forall\alpha \in[0,1]$ $\tilde{y}_\alpha^- \leq \widetilde{m}_\alpha^-$ and $\widetilde{m}_\alpha^+ \leq \tilde{y}_\alpha^+$.

Proof:

Let $\widetilde{m}_\alpha = \cap_{\tilde{y}\in<\widetilde{x}>}[\tilde{y}_\alpha^-,\tilde{y}_\alpha^+]$, then $\widetilde{m}_\alpha$ is closed and convex and therefore an interval. If $\beta > \alpha$ then $[\widetilde{y}_\beta^-, \widetilde{y}_\beta^+] \subseteq [\widetilde{y}_\alpha^-, \widetilde{y}_\alpha^+]$, thus

$$\bigcap_{\widetilde{y}\in<\widetilde{x}>} [\widetilde{y}_\beta^-, \widetilde{y}_\beta^+] \subseteq \bigcap_{\widetilde{y}\in<\widetilde{x}>} \left[\widetilde{y}_\alpha^-, \widetilde{y}_\alpha^+\right]$$

and $\widetilde{m}$ is a fuzzy number with the desired property.□

Theorem 25 (Diamond [7]) $\mathcal{L}$ with addition and scalar multiplication of fuzzy numbers forms a vector space over R.

Recall that for a given possibility distribution for the random variable

$\tilde{x}$, the function EA($\tilde{x}$) of Definition 16 provides an estimate of the expected value of $\tilde{x}$. Note that for all members of a given equivalence class $<\tilde{x}>\in \mathcal{L}$, this functional will result in the same estimate. Since $EA(\tilde{x})$ determines the midpoint of the interval containing the actual expected value of $\tilde{x}$ when $\tilde{x}$ is a random variable, this functional is consistent with the assumed utility of the decision maker for an interval of possible values. This motivates the following extension of this definition to $\mathcal{L}$.

Definition 23 Let $<\tilde{x}>\in \mathcal{L}$ the **expected average** of $<\tilde{x}>$ is

$$EA(<\tilde{x}>) = \frac{1}{2}\int_0^1 (\tilde{x}_\alpha^- + \tilde{x}_\alpha^+)d\alpha. \tag{3.2}$$

where $\tilde{x} \in <\tilde{x}>$.

Example 13 Let $\tilde{x}$ be the fuzzy number with set representation $[1+\alpha, 3-\alpha]_{\alpha\in[0,1]}$, then $EA(<\tilde{x}>) = \frac{1}{2}\int_0^1 (1+\alpha+3-\alpha)d\alpha = \int_0^1 2d\alpha = 2$. This is not surprising since the midpoint of each $\alpha-$cut is 2.

Example 14 Let $\tilde{x}$ be the fuzzy number with set representation $[2+\alpha, 4-\alpha^2]_{\alpha\in[0,1]}$. Then $EA(<\tilde{x}>) = \int_0^1 \frac{1}{2}(6+\alpha-\alpha^2)d\alpha. = 3.0833$ The expected average is to the right of 3, since the fuzzy number is weighted to the right of 3.

Example 15 Let $\tilde{x}$ be the fuzzy number with set representation equal [1,2] for $\alpha = (\frac{1}{2}, 1]$ and [0,2] for $\alpha = [0, \frac{1}{2}]$. Then $EA(<\tilde{x}>) = \int_{\frac{1}{2}}^1 1.5d\alpha + \int_0^{\frac{1}{2}} d\alpha =$

$\frac{1}{2}(1.5) + \frac{1}{2}(1) = 1.25.$

It has been shown that $\mathcal{L}$ is a vector space. A norm can be put on this space using the expected utility concept.

Definition 24 On the vector space $\mathcal{L}$, let the function

$$\| < \tilde{x} > \|_{EA} = \int_0^1 \frac{1}{2} \left| \tilde{x}_\alpha^- + \tilde{x}_\alpha^+ \right| d\alpha \tag{3.3}$$

where $\tilde{x} \in < \tilde{x} >$.

This is well defined since $\tilde{x}_\alpha^-$ and $\tilde{x}_\alpha^+$ are integrable on [0,1] since each is monotonic (see 6.9 in Kolmogorov&Fomin [24]). This implies that $\tilde{x}_\alpha^- + \tilde{x}_\alpha^+$ is integrable which implies that $|\tilde{x}_\alpha^- + \tilde{x}_\alpha^+|$ is integrable. $\| < \tilde{x} > \|_{EA}$ measures the expected absolute value of the midpoint of each α-cut. As mentioned earlier, the subscript "EA" is an abbreviation for *expected average.* A fuzzy number that has membership value zero, at all but a single point where the membership value is one, is called a crisp number. For a crisp number the expected average norm reduces to the absolute value of the number.

Example 16 Consider the triangular fuzzy number $\tilde{x}$=(-2,-1,1), i.e. it's $\alpha - cuts$ are given by $\tilde{x}_\alpha = [-2 + \alpha, 1 - 2\alpha]$. Then

$$\| < \tilde{x} > \|_{EA} = \int_0^1 \frac{1}{2} \left| -2 + \alpha + 1 - 2\alpha \right| d\alpha = .75.$$

Next it will be shown that this function defines a norm on $\mathcal{L}$. To do this, the following lemma is needed.

Lemma 1 (Kruse et al. [23]) Let $\tilde{x} \in \mathcal{E}$. Consider $\tilde{x}_\alpha^-$ and $\tilde{x}_\alpha^+$ as functions of α. Then $\tilde{x}_\alpha^-$ and $\tilde{x}_\alpha^+$ are continuous from the left.

Observe that these functions are monotonic on [0,1] and so are continuous except at countably many points in [0,1] and all discontinuities are of the first kind (see Rudin [39] page 96).

Theorem 26 The function $\| < \tilde{x} > \|_{EA}$ defines a norm on $\mathcal{L}$.

Proof:

(N1) $\| < \tilde{x} > \|_{EA} \geq 0$ since $|\tilde{x}_\alpha^- + \tilde{x}_\alpha^+| \geq 0 \ \forall \alpha \in [0,1]$

(N2) $\| < \tilde{x} > \|_{EA} = 0 \Rightarrow \tilde{x} = 0$.

Case(1) Assume for some $\gamma > 0$ we have $\left|\tilde{x}_\gamma^- + \tilde{x}_\gamma^+\right| = r > 0$. Since $\tilde{x}_\beta^- \to \tilde{x}_\gamma^-$ as $\beta \to \gamma$ from below and similarly for $\tilde{x}_\gamma^+$, $\exists \delta > 0$ such that $\left|\tilde{x}_\beta^- - \tilde{x}_\gamma^-\right| < r/3$ and $\left|\tilde{x}_\beta^+ - \tilde{x}_\gamma^+\right| < r/3$ and thus $\left|\tilde{x}_\beta^- + \tilde{x}_\beta^+\right| > 0 \forall \beta < \gamma$ with $\gamma - \beta < \delta$.

Then $\int_{\gamma-\delta}^{\gamma} \frac{1}{2} |\tilde{x}_\alpha^- + \tilde{x}_\alpha^+| \, d\alpha > 0 \Rightarrow \int_0^1 \frac{1}{2} |\tilde{x}_\alpha^- + \tilde{x}_\alpha^+| \, d\alpha > 0$, a contradiction. Thus $\forall \alpha > 0, |\tilde{x}_\alpha^- + \tilde{x}_\alpha^+| = 0 \Rightarrow \tilde{x}_\alpha^- = -\tilde{x}_\alpha^+$ i.e. $\tilde{x} \equiv 0$.

Case(2) Let $|\tilde{x}_0^- + \tilde{x}_0^+| = r > 0$. Then for some γ arbitrarily close to 0 and $r > s > 0$

$\left|\tilde{x}_\gamma^- + \tilde{x}_\gamma^+\right| = s > 0$ and case(1) applies.

(N3) $\|r < \tilde{x} > \|_{EA} = |\mathrm{r}| \, \| < \tilde{x} > \|_{EA}$. Clear since absolute value is a norm.

(N4)

$$\|< \tilde{x} + \tilde{y} >\|_{EA} \leq \|< \tilde{x} >\|_{EA} + \|< \tilde{y} >\|_{EA}\,.$$

$$\|< \tilde{x} + \tilde{y} >\|_{EA} = \int_0^1 \tfrac{1}{2}\,|\tilde{x}_\alpha^- + \tilde{y}_\alpha^- + \tilde{x}_\alpha^+ + \tilde{y}_\alpha^+|\,d\alpha$$

$$\leq \int_0^1 \tfrac{1}{2}\,\{|\tilde{x}_\alpha^- + \tilde{x}_\alpha^+| + |\tilde{y}_\alpha^- + \tilde{y}_\alpha^+|\}\,d\alpha = \|< \tilde{x} >\|_{EA} + \|< \tilde{x} >\|_{EA}$$

since absolute value is a norm and by properties of the Lebesgue integral. □

3.3 An Isometry between $(\mathcal{L}, \|\cdot\|_{\mathcal{EA}})$ and BV[0,1] as a Subspace of $\mathbf{L}_1[0,1]$

For this section, the following definitions and results from the theory of functions of bounded variations are needed.

Definition 25 (Kolmogorov&Fomin [24]) Let f:[a,b]→ R. f is said to be **of bounded variation** if $\exists C > 0$ such that $\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \leq C$ for every partition $a = x_0 < x_1 < ... < x_k = b$ on [a,b]. We denote the set of all functions of bounded variation on [a,b] by **BV[a,b]**.

Definition 26 (Royden [38]) Let f:[a,b]→ R be a function of bounded variation. The **total, positive and negative variation** of f on [a,b] are denoted

$$\mathbf{V}_a^b(\mathbf{f}) = sup_p \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|,$$

$$\mathbf{P}_a^b(\mathbf{f}) = \mathrm{sup}_p \sum_{i=1}^{n} \max\{0, f(x_i) - f(x_{i-1})\},$$

$$\mathbf{N}_a^b(\mathbf{f}) = \mathrm{sup}_p \sum_{i=1}^{n} \max\{0, f(x_{i-1}) - f(x_i)\}$$

where p represents all partitions of [a,b].

Theorem 27 (Royden [38]) Let f:[a,b]→R be a function of bounded variation. Then $\mathbf{V}_a^b\mathbf{(f)}$=$\mathbf{P}_a^b\mathbf{(f)}$+$\mathbf{N}_a^b\mathbf{(f)}$ and f(b)-f(a)=$\mathbf{P}_a^b\mathbf{(f)}$-$\mathbf{N}_a^b\mathbf{(f)}$.

Theorem 28 (Kolmogorov&Fomin [24]) Let $f : [a, b] \to R$ be a function of bounded variation and $a < b < c$. Then $\mathbf{V}_a^c\mathbf{(f)}$=$\mathbf{V}_a^b\mathbf{(f)}$+$\mathbf{V}_b^c\ \mathbf{(f)}$.

It is now shown that each element of $\mathcal{L}$ can be represented by an equivalence class of functions from BV[0,1], the space of bounded variations on the interval [0,1] when BV[0,1] is considered a subspace of $L_1[0,1]$ (the space of Lebesgue integrable functions on [0,1] partitioned into functions which are equivalent up to a set of measure zero). Using this representation, it will be shown that $\mathcal{L}$ is isometric to this space under the EA-norm above.

Theorem 29 (Cadenas&Verdegay [5]) Let $(A_\alpha)_{\alpha\in[0,1]}$ be a family of non-empty closed nested intervals with $\alpha < \beta \Rightarrow A_\beta \subseteq A_\alpha$ and $A_0 = \text{cls}(\cup_{\gamma\in(0,1]}A_\gamma)$. Define $\mu_{\tilde{x}}(x)$ =sup$\{\alpha \mid x \in A_\alpha\}$ with sup$\emptyset \equiv 0$. Then $\tilde{x}$ is a fuzzy number.

Definition 27 Let $\tilde{x}$ be a fuzzy number and let $(A_\alpha)_{\alpha\in[0,1]}$ be a set representation of $\tilde{x}$ consisting of non-empty closed intervals with

$$A_0 = \text{cls}(\bigcup_{\gamma\in(0,1)} A_\gamma).$$

Define $f_{\tilde{x}}$:[0,1]→R, called a **functional representative of** $\tilde{x}$, by $f_{\tilde{x}}(\alpha) = (a_\alpha^- + a_\alpha^+)/2$ where $A_\alpha=[a_\alpha^-,a_\alpha^+]$.

Theorem 30 Let $f_{\tilde{x}}$ be a functional representative of x̃, then $f_{\tilde{x}}$ $\in$BV[0,1] and $f_{\tilde{x}}(\alpha\text{-})=(\beta^-+\beta^+)/2$ where $[\beta^-,\beta^+]=\cap_{\gamma<\alpha}A_\gamma=\tilde{x}_\alpha$ and $f_{\tilde{x}}(\alpha+)=(\beta^-+\beta^+)/2$ where $[\beta^-,\beta^+]=\mathrm{cls}(\cup_{\gamma>\alpha}A_\gamma)$.

Proof:

Note that $f_{\tilde{x}}(\alpha) = (a_\alpha^- + a_\alpha^+)/2 = (f_{\tilde{X}}^- - (-f_{\tilde{X}}^+))/2$ where $f_{\tilde{X}}^-(\alpha) = a_\alpha^-$ and $f_{\tilde{X}}^+(\alpha) = -a_\alpha^+$. Since $(A_\alpha)_{\alpha\in[0,1]}$ is nested and $f_{\tilde{X}}^-$ and $-f_{\tilde{X}}^+$ are monotone increasing real-valued functions, $f_{\tilde{x}}$ is a bounded variation (see Theorem 4. on page 100 of Royden [38]). Also, $f_{\tilde{x}}(\alpha\text{-})$ and $f_{\tilde{x}}(\alpha+)$ exist (since the left and right limits exist for monotonic functions, see Theorem 4.29 on page 95 of Rudin [39]). Let $[\beta^-,\beta^+]=\cap_{\gamma<\alpha}A_\gamma$. We know that $f_{\tilde{x}}^-(\alpha-) = \beta^-$ and $-f_{\tilde{x}}^+(\alpha-) = \beta^+$ since each is monotone increasing. Thus, $f_{\tilde{x}}(\alpha-) = (f_{\tilde{x}}^-(\alpha-) - (-f_{\tilde{x}}^+(\alpha-)))/2 = (\beta^- + \beta^+)/2$. Similarly for the limit from the right.$\Box$

Definition 28 Let f,g$\in$BV[0,1] and let f$\sim$g mean that f(x)=g(x) almost everywhere. This defines an equivalence relation over BV[0,1]. Let $\mathrm{BVL}_1[0,1]$ denote BV[0,1] for this partition with respect to the L_1 norm.

Theorem 31 Let $(A_\alpha)_{\alpha\in[0,1]}$ and $(B_\alpha)_{\alpha\in[0,1]}$ be set representations of fuzzy number, x̃ consisting of non-empty closed intervals. Moreover, let f and g be the corresponding functional representatives. Then f$\sim$g.

Proof:

Since $\tilde{x}_\alpha = \cap_{\gamma<\alpha} A_\gamma = \cap_{\gamma<\alpha} B_\gamma$ $\forall\alpha \in (0,1]$ we have $f(\alpha-)=g(\alpha-)$. But this implies that $f \sim g$ since the points at which a function of bounded variation are discontinuous are countable and the measure of a countable set is zero. At all points of continuity we have $f(\alpha)=f(\alpha-)=g(\alpha-)=g(\alpha)$.$\square$

Theorem 32 Let $(A_\alpha)_{\alpha\in[0,1]}$ and $(B_\alpha)_{\alpha\in[0,1]}$ be set representations of fuzzy numbers $\tilde{x}$ and $\tilde{y}$ consisting of non-empty closed intervals and assume $\tilde{x}\sim\tilde{y}$, then if f and g are functional representations for these two set representations respectively, $f\sim g$.

Proof:

$\forall\alpha \in[0,1]$, $(a_\alpha^- + a_\alpha^+)/2 = (b_\alpha^- + b_\alpha^+)/2$ where $\tilde{x}_\alpha = [a_\alpha^-, a_\alpha^+]$ and $\tilde{y}_\alpha = [b_\alpha^-, b_\alpha^+]$ by definition of our equivalence relation. Thus $f(\alpha-)=g(\alpha-)$ and the same argument as applied in the previous theorem holds. $\square$

Theorems 30, 31 and 32 show that there is a single functional representation (up to a set of measure zero) for each equivalence class of fuzzy numbers. Thus, the map of fuzzy number equivalence classes to the space of equivalence classes of functions of bounded variation is well defined. The next step is to show that it is not only well-defined but establishes an isomorphism. The next theorem establishes that the map is one-to-one and the following one establishes that it is onto.

Theorem 33 Let $< \tilde{x} > \neq < \tilde{y} >$ be two fuzzy number equivalence classes. Then their functional representations are not equal.

Proof:

Let f and g be the functional representations for $< \tilde{x} >$ and $< \tilde{y} >$ respectively. Since $< \tilde{x} > \neq < \tilde{y} >$ $\exists \alpha \in (0,1]$ such that $(\tilde{x}_\alpha^- + \tilde{x}_\alpha^+)/2 \neq (\tilde{y}_\alpha^- + \tilde{y}_\alpha^+)/2$. Then for this α, $f(\alpha-) \neq g(\alpha-)$. Then $\exists$ disjoint neighborhoods of $f(\alpha-)$ and $g(\alpha-)$ with positive separation so in a neighborhood near α (a set with measure greater than zero), $f \neq g$.□

Theorem 34 Let $f \in BV[0,1]$ then $\exists < \tilde{x} > \in \mathcal{L}$ such that f is the functional representation for $< \tilde{x} >$.

Proof:

For $\alpha \in (0,1]$, let $[\tilde{x}_\alpha^-, \tilde{x}_\alpha^+] = [f(1) - 2\mathbf{P}_a^1(\mathbf{f}), f(1) + 2\mathbf{N}_a^1(\mathbf{f})]$ and $[\tilde{x}_0^-, \tilde{x}_0^+] = \text{cls}\,(\bigcup_{\alpha>0} [\tilde{x}_\alpha^-, \tilde{x}_\alpha^+])$. The claim is that the intervals $[\tilde{x}_\alpha^-, \tilde{x}_\alpha^+]$ are a set representation for a fuzzy number $\tilde{x}$. Since each interval is closed and non-empty, by our prior theorem we only need to show that $\alpha < \beta \Rightarrow [\tilde{x}_\beta^-, \tilde{x}_\beta^+] \subset [\tilde{x}_\alpha^-, \tilde{x}_\alpha^+]$, but this is immediate from the definition of $\mathbf{P}_a^1(\mathbf{f})$ and $\mathbf{N}_a^1(\mathbf{f})$.

To show that this is a functional representation for $< \tilde{x} >$ note that

$g(\alpha) = (\tilde{x}_\alpha^- + \tilde{x}_\alpha^+)/2 = f(1) + (2\mathbf{N}_a^1(\mathbf{f}) - 2\mathbf{P}_a^1(f))/2 = f(1) + (f(\alpha) - f(1)) = f(\alpha)$□

Theorem 35 The space, $\mathcal{L}$ is isomorphic to $\mathrm{BVL}_1[0,1]$.

Proof:

Theorems 33 and 34 establish a one-to-one correspondence between $\mathcal{L}$ and $\mathrm{BVL}_1[0,1]$. If $<\tilde{x}>, <\tilde{y}> \in \mathcal{L}$ with functional representations f and g, then f+g is a functional represenation for $<\tilde{x}> + <\tilde{y}>$. To see this let $(\mathrm{A}_\alpha)_{\alpha\in[0,1]}$ and $(\mathrm{B}_\alpha)_{\alpha\in[0,1]}$ be the set representations of $\tilde{x}$ and $\tilde{y}$ corresponding to f and g. We know by Theorem 13 that $(\mathrm{A}_\alpha+B_\alpha)_{\alpha\in[0,1]}$ is a set representation for $\tilde{x}+\tilde{y}$. But the mipoint of $\mathrm{A}_\alpha+B_\alpha$ is just the sum of the midpoints of A_α and B_α. So f+g is a functional representation for $\tilde{x}+\tilde{y}$. Similar reasoning shows that cf is a functional representation for $c\tilde{x}$ for scalor c. Thus the mapping preserves addition and multiplication by scalars and is an isomorphism.$\Box$

Theorem 36 The space $(\mathcal{L}, \|\cdot\|_{\mathcal{EA}})$ is isometric to $\mathrm{BVL}_1[0,1]$.

Proof:

Since an isomorphism between $\mathcal{L}$ and $\mathrm{BVL}_1[0,1]$ has already been established in Theorem 35, we need only to establish an equivalent norm. Note that if $<\tilde{x}> \in \mathcal{L}$ and $f_{\tilde{x}}$ a the functional representation for $<\tilde{x}>$ then $f_{\tilde{x}} \sim g_{\tilde{x}}$ where $g_{\tilde{x}}$ is the functional representation for $<\tilde{x}>$ formed by the α-cuts of $\tilde{x}$. But then $\|<\tilde{x}>\|_{EA} = \int_0^1 |g_{\tilde{x}}(\alpha)|\, d\alpha = \int_0^1 |f_{\tilde{x}}(\alpha)|\, d\alpha$. Thus the norm is independent of the choice of functional representation since we have shown in

Theorem 31 that all functional representations for $< \widetilde{x} >$ are equal almost everywhere. But $f_{\widetilde{x}}$ is an element of BVL$_1$[0,1] as a subspace of L$_1[0,1]$ and in L$_1[0,1]$ $\|f_{\tilde{x}}\|_1 = \int_0^1 |\mathrm{f}_{\tilde{x}}(\alpha)|\, d\alpha$. Therefore, the norms are equivalent.□

3.4 Convergence in $(\mathcal{L}, \|\cdot\|_{\mathcal{EA}})$

It has been established that the equivalence classes of fuzzy numbers are isometric to the space BVL$_1$[0,1] which is a subspace of L$_1$[0,1]. Thus every Cauchy sequence of fuzzy number equivalence classes in the space $\mathcal{L}$ will converge but not necessarily to a fuzzy number equivalence class. However, we have the following result that shows that Cauchy sequences in the space converge if our fuzzy numbers have a uniform bound in variation.

Theorem 37 Suppose $\{< \widetilde{x} >_n\}$ is Cauchy in $\mathcal{L}$ under the EA-norm and suppose that $\exists$ M$\in$ R such that $\forall$ n, $\mathrm{V}_0^1(f_n) < M$, where f$_n$ is a functional representation for $< \widetilde{x} >_n$. Then $\exists < \widetilde{x} > \in \mathcal{L}$ such that $< \widetilde{x} >_n \rightarrow < \widetilde{x} >$ in the EA-norm.

Proof:

By assumption $\{f_n(\alpha)\}$ is Cauchy in L$_1[0,1]$. Thus $\exists$ f$\in$ L$_1[0,1]$ such that $f_n(\alpha) \rightarrow f(\alpha)$ in the mean since L$_1[0,1]$ is complete. We know that $\exists$ a subsequence $f_{n_k}(\alpha) \rightarrow f(\alpha)$ almost everywhere (see Kolmogorov&Fomin [24] page 388 problem 7c). Let S equal the subset of [0,1] where we have pointwise convergence. Assume that f is not of bounded variation on S. Since f is not of bounded variation $\exists$ $\{\alpha_i \in S \mid i = 1, m\}$ such that $\sum \left| f(\alpha_i) - f(\alpha_{i+1)} \right| > 2M$. We can chose N large enough such that $\forall i$, $|f(\alpha_i) - f_N(\alpha_i)| < \epsilon$ for ϵ arbitrarily small. Then $\sum \left| f_N(\alpha_i) - f_N(\alpha_{i+1)} \right| > \sum \left| f(\alpha_i) - f(\alpha_{i+1)} \right| - m\epsilon$. But then $\mathrm{V}_0^1(f_N) > M$ which is a contradiction. Therefore f is of bounded variation on S. We need to show that there is a function of bounded variation on [0,1] which is equal to f almost everywhere. For every $\alpha \in [0,1] - S$. Let $\{x_n\}$ be a sequence from S such that $\mathrm{x}_n \to \alpha$. Such a sequence exists since [0,1]-S has measure zero. Define f$(\alpha) = \limsup f(x_n)$. This limit exists and is finite since f is a bounded variation on S and, therefore, bounded on S. But then f as just defined is a bounded variation on [0,1] since $\forall \alpha \in [0,1] - S$ we can find $x \in S$ arbitrarily close to α such that $|f(x) - f(\alpha)| < \epsilon$ for arbitrarily small ϵ . $\square$

Corollary 4 Let $\{\tilde{x}_n\}$ be a sequence of fuzzy numbers with the property that $\forall n$ $(\tilde{x}_n)_0^+ - (\tilde{x}_n)_0^- < 2M$. If $\{< \tilde{x} >_n\}$, the sequence of equivalence classes in $\mathcal{L}$, is Cauchy then it converges.

Proof:

Let $\mathrm{f}_{\tilde{x}_n}$ be defined by $\mathrm{f}_{\tilde{x}_n}(\alpha)$=$((\tilde{x}_n)_\alpha^- + (\tilde{x}_n)_\alpha^+)/2$ where $(\tilde{x}_n)_\alpha$=$[(\tilde{x}_n)_\alpha^-, (\tilde{x}_n)_\alpha^+]$. Thus $\mathrm{f}_{\tilde{x}_n}$ is the functional representative of $< \tilde{x}_n >$ formed by the $\alpha - cuts$ of $\tilde{x}_n$. Applying the theorem, we will be finished if we show that $2\mathrm{V}_0^1(f_{\tilde{x}_n}) \leq (\tilde{x}_n)_0^+ - (\tilde{x}_n)_0^- < 2M$.

For each n we define the fuzzy number $\widetilde{m}_n$ with set representation given by $(\widetilde{m}_n)_\alpha$ =$[(\widetilde{m}_n)_\alpha^-, (\widetilde{m}_n)_\alpha^+]$=[$f_{\widetilde{x}_n}(1)-2\mathbf{P}_\alpha^1(\mathbf{f}_{\widetilde{x}_n})$,$f_{\widetilde{x}_n}(1)$+2$\mathbf{N}_\alpha^1(\mathbf{f}_{\widetilde{x}_n})$] $\forall \alpha \in (0,1]$ and $[(\widetilde{m}_n)_0^-, (\widetilde{m}_n)_0^+]$=cls($\cup_{\alpha>0}[(\widetilde{m}_n)_\alpha^-, (\widetilde{m}_n)_\alpha^+]$). We have shown that $\widetilde{m}_n \in < \tilde{x}_n >$. We claim that $\widetilde{m}_n$ is the element of $< \tilde{x}_n >$ with minimal possibility. Assume not. Then $\exists \tilde{y} \in < \tilde{x}_n >$ such that for some $\alpha \in [0,1]$, $\tilde{y}_\alpha^+ < (\widetilde{m}_n)_\alpha^+$ and $(\widetilde{m}_n)_\alpha^- < \tilde{y}_\alpha^-$ i.e. $\tilde{y}_\alpha^+ - \tilde{y}_\alpha^- < (\widetilde{m}_n)_\alpha^+ - (\widetilde{m}_n)_\alpha^-$ $=2\mathbf{V}_\alpha^1(\mathbf{f}_{\widetilde{x}_n})$. But we know that $f_{\widetilde{x}_n}(\alpha) = \frac{\tilde{y}_\alpha^+}{2} + \frac{\tilde{y}_\alpha^-}{2}$ which implies that $2\mathbf{V}_\alpha^1(\mathbf{f}_{\widetilde{x}_n}) \leq \mathbf{V}_\alpha^1(\tilde{\mathbf{y}}_\alpha^+) + \mathbf{V}_\alpha^1(\tilde{\mathbf{y}}_\alpha^-)$. Since the functions on the right-hand side are monotonic, their variation is defined by their values at the end points. Thus $2\mathbf{V}_\alpha^1(\mathbf{f}_{\widetilde{x}_n}) \leq \tilde{y}_\alpha^+ - \tilde{y}_1^+ + \tilde{y}_1^- - \tilde{y}_\alpha^- \leq \tilde{y}_\alpha^+ - \tilde{y}_\alpha^-$ and we have arrived at a contradiction.

Since $\widetilde{m}_n$ is the distribution of minimal possibility for $< \tilde{x}_n >$, $2\mathbf{V}_\alpha^1$ $(\mathbf{f}_{\widetilde{x}_n})$= $(\widetilde{m}_n)_0^+ - (\widetilde{m}_n)_0^- \leq (\tilde{x}_n)_0^+ - (\tilde{x}_n)_0^- < 2M$. □

Corollary 5 Let $\{\tilde{x}_n\}$ be a sequence of fuzzy numbers with the property that $\forall n$ $(\tilde{x}_n)_0 \subset B$ where B is a bounded subset of R and $(\tilde{x}_n)_0$ is the support of $\tilde{x}_n$. If $\{< \tilde{x} >_n\}$, the sequence of equivalence classes in $\mathcal{L}$, is Cauchy then it converges.

Proof:

This follows from the prior corollary since $(\tilde{x}_n)_0 \subset B \Rightarrow (\tilde{x}_n)_0^+ - (\tilde{x}_n)_0^- < 2M$ where M is an upper bound on the absolute values of the elements of B.□

4. Optimization of Fuzzy Functions

We begin this chapter with a definition of a fuzzy function. We show that fuzzy functions as so defined are mappings into the space of fuzzy vectors (also defined in this chapter). In particular, real valued fuzzy functions are mappings to the space of fuzzy numbers. We will examine unconstrained convex fuzzy functions. The minimum of a fuzzy function and the minimizer of a fuzzy function will be defined and their properties examined. We will consider the concept of minimizing a fuzzy function in the context of possibility theory. This allows formulation of optimization problems involving real valued fuzzy functions in the space $(\mathcal{L}, \|\cdot\|_{\mathcal{E}\mathcal{A}})$.

4.1 Fuzzy Functions

In this section a definition of a fuzzy function as a special possibility distribution over a function space is introduced. Attention is focused upon functions of real finite dimensional vector spaces. A definition of a fuzzy vector which is a slight generalization of the definition of a fuzzy number is given first.

Definition 29 Let X be a real finite dimensional vector space with the Euclidean norm. A **fuzzy vector**, $\tilde{\mathrm{x}}$, is a fuzzy subset of X such that 1) $\tilde{\mathrm{x}}$ is normal 2) $\forall\alpha \in (0,1]$ $\tilde{x}_\alpha$ is compact and 3) $\tilde{x}_{0+}$ is bounded. A **convex fuzzy vector**, $\tilde{x}$, is a fuzzy vector such that $\forall\alpha \in (0,1]$ $\tilde{x}_\alpha$ is convex.

In Lee [25] a convex fuzzy vector is referred to as an approximate quantity. This definition of fuzzy vector is consistent with the definition of fuzzy vector given in Buckley [3]. If X=R, the definition of a convex fuzzy vector, the definition of approximate quantity in Lee [25] and fuzzy number in Buckley [3], Kaufmann&Gupta [20] and Zhang [50] all coincide.

Example 17 A vector of fuzzy numbers, $(\tilde{a}_i)$, $i = 1, \cdots, n$, is a convex fuzzy vector. Each α-cut is $\Pi_{i=1,n}[(\tilde{a}_i)^-_\alpha, (\tilde{a}_i)^+_\alpha]$ where $\tilde{a}_\alpha$=$[(\tilde{a}_i)^-_\alpha, (\tilde{a}_i)^+_\alpha]$. But not all convex fuzzy vectors can be so represented. For example, consider the convex fuzzy vector with α-cuts, $\tilde{x}_\alpha = \{x \mid \|x - x_0\| \leq 1 - \alpha\}$, where $\alpha \in (0,1]$.

Definition 30 Let R^n,R^m be real finite dimensional vector spaces with the Euclideanl norm, $\|\cdot\|$ and $\Omega \subset R^n$,

$$\mathcal{F} = \{\text{f: } \Omega \to R^m \mid \text{f is a bounded function over } \Omega\}.$$

On $\mathcal{F}$ we define the norm $\|f\|_{\sup}$=$\sup_{x\in\Omega} \|f(x)\|$ (see Rudin [39]-def.7.14 and note continuity is not used in the norm proof). Let $\tilde{f}$ be a fuzzy subset of $\mathcal{F}$. For x$\in$ Ω define $\tilde{f}$(x) to be the fuzzy subset of R^m with membership function $\mu_{\tilde{f}(x)}$(y)=$\sup\{\alpha : \mu_{\tilde{f}}(f) = \alpha,\ f \in \mathcal{F} \text{ and } f(x) = y\}$. A **fuzzy function** over Ω is a fuzzy subset $\tilde{f}$ of $\mathcal{F}$ such that 1)$\tilde{f}$ is normal 2) $\forall\alpha \in$ (0,1] $\tilde{f}_\alpha$ is path connected and compact and 3) $\tilde{f}_{0+}$ is bounded. If $\forall\alpha \in (0,1]$, $\tilde{\mathrm{f}}_\alpha$ is convex we say $\tilde{\mathrm{f}}$ is a **convex fuzzy function**.

Write $\tilde{\mathrm{f}}$: $\Omega \rightarrow \mathcal{E}^m$ where $\mathcal{E}^m$ is the set of all fuzzy vectors over R^m. In a moment it will be shown that this notation makes sense.

This definition of fuzzy function is consistent with the functions which define the linear programing problems examined in Buckley [3], Lodwick [27] and Jamison&Lodwick [15]. In these papers, the functions are defined in terms of linear equations in coefficients which are fuzzy numbers. This definition is not equivalent to the definition of fuzzy mapping used in Lee [25] and Zhang [50]. A fuzzy mapping is defined as a point to fuzzy set mapping.

A fuzzy function, as defined here, allows for more refinement in the definition by restricting the functions in $\tilde{\mathrm{f}}_\alpha$. For example, the common properties that crisp elements of $\tilde{\mathrm{f}}$ possess can be used to characterize $\tilde{\mathrm{f}}$ as in the following definitions.

Definition 31 Let $\tilde{\mathrm{f}}$:$\Omega \rightarrow \mathcal{E}$. Call $\tilde{\mathrm{f}}$ a **fuzzy convex function** if $\forall\alpha \in [0,1]$ and $\forall \mathrm{f} \in \tilde{\mathrm{f}}_\alpha$, f is a convex function.

Definition 32 Let $\tilde{\mathrm{f}}$:$\Omega \rightarrow \mathcal{E}$. Call $\tilde{\mathrm{f}}$ a **fuzzy continuous function** if $\forall\alpha \in [0,1]$ and $\forall \mathrm{f} \in \tilde{\mathrm{f}}_\alpha$, f is a continuous function.

Remark 3. To say $\tilde{f}_\alpha$ is **path connected** means that $\forall f_1$ and $f_2 \in \tilde{f}_\alpha$ $\exists$ a continuous function (under the supremum norm) $F:[0,1] \rightarrow \mathcal{F}$ such that $F(0)=f_1$ and $F(1)=f_2$.

Remark 4. For a fuzzy function to be a convex fuzzy function an equivalent statement is: $\forall \alpha, \beta \in [0,1]$, f_1 and $f_2 \in \tilde{f}_\alpha$ and $x \in \Omega$, $\exists f_3 \in \tilde{f}_\alpha$ such that $f_3(x)=\beta f_1(x)+(1-\beta)f_2(x)$.

Example 18 One way to write a fuzzy function is to write it as a function of fuzzy parameters. A parameterized fuzzy function of this type is denoted by $f(\tilde{a},x)$ where x is a vector in R^n and $\tilde{a}$ is a vector whose entries are fuzzy numbers. For example, let $\tilde{a}$ be the triangular fuzzy number defined by the parameters (1,2,3). Define a fuzzy function $\tilde{f}(x)=f(\tilde{a},x)=\tilde{a}^*x^2$ with x restricted to a compact set in R. More formally, $\mu_{\tilde{f}}(f)=\alpha \Rightarrow f(x)=a^*x^2$ where $\mu_{\tilde{a}}(a)=\alpha$. The range of values of $\tilde{f}_{.5}(x)$ are between and including $f(x) = 1.5 * x^2$ and $g(x) = 2.5 * x^2$ with both f and $g \in \tilde{f}_{.5}$. Hence $\tilde{f}_{.5}(1)=[\frac{3}{2},\frac{5}{2}]$.

The following theorem demonstrates that the definition of a fuzzy function produces the desired result, namely, that it is a point to fuzzy vector mapping.

Theorem 38 Let $\tilde{f}$: $\Omega \to \mathcal{E}^n$ be a fuzzy function. Then $\tilde{f}(x)$ is a fuzzy vector $\forall x \in \Omega$. If $\tilde{f}$ is a convex fuzzy function, then $\tilde{f}(x)$ is a convex fuzzy vector.

Proof:

Assume $\tilde{f}$ is a fuzzy function.

(Normality) $\tilde{f}$ normal $\Rightarrow \exists f \in \tilde{f}_1 \Rightarrow \forall x \in X$, $y=f(x) \in \tilde{f}_1(x) \Rightarrow \tilde{f}(x)$ normal.

(Compactness) Let $(f_n(x)) \subset \tilde{f}(x)_\alpha$ where $(f_n) \subset \tilde{f}_\alpha$. $\tilde{f}_\alpha$ compact $\Rightarrow \exists$ $(f_{n_i}) \subset \tilde{f}_\alpha$ such that $f_{n_i} \to f \in \tilde{f}_\alpha$. Hence $\|f_{n_i} - f\|_{\sup} \to 0 \Rightarrow f_{n_i}(x) \to f(x) \in \tilde{f}_\alpha(x)$.

Hence $\tilde{f}(x)$ is a fuzzy vector. If $\tilde{f}$ is a convex fuzzy function then $\tilde{f}(x)_\alpha$ is convex by definition.□

It was shown in Example 18, that a function with fuzzy parameters can be a fuzzy function. But this is not always true. The following two theorems examine the limits of these representations.

Theorem 39 Let $f(\tilde{a},x):\Omega \to \mathcal{E}$ represent the fuzzy subset of functions given by $\mu_{f(\tilde{a},x)}(f(a,x)) = \mu_{\tilde{a}}(a)$, where $x \in \Omega \subset R^n$ and $\tilde{a}$ is a vector of fuzzy numbers, that is $\tilde{a}=[\tilde{a}_1,\tilde{a}_2,\ldots,\tilde{a}_n]^T$, $\tilde{a}_i$ a fuzzy number. Then for each $\alpha \in [0,1]$: $\tilde{f}_\alpha = \{f(a,x):\Pi_{i=1,n}[(\tilde{a}_i)^-_\alpha, (\tilde{a}_i)^+_\alpha] \times \Omega \to R\}$. For each $f(a,x)$, define the functions $f_a(x)=f(a,x)$ and $f_x(a)=f(a,x)$ where a and x are held fixed respectively. Assume that $\forall \alpha \in [0,1]$ and $x \in \Omega$, f_x is continuous over $\tilde{a}_\alpha = \Pi_{i=1,n}[(\tilde{a}_i)^-_\alpha, (\tilde{a}_i)^+_\alpha]$. Then $f(\tilde{a},x)$ is a convex fuzzy function if and only if $\{f_x:\tilde{a}_\alpha \to R \mid x \in \Omega\}$ is an equicontinuous family of functions and each f_a is bounded over Ω.

Proof:

$\Rightarrow$ (Equicontinuity) Let $\alpha \in [0,1]$ and assume $\{f_x:\tilde{a}_\alpha \to R \mid x \in \Omega\}$ is not equicontinuous. Then $\exists\epsilon > 0$ such that $\forall\ n\exists x_n, a_n$ and b_n with $|a_n - b_n| < \frac{1}{n}$ but $\left|f_{x_n}(a_n)\text{-}f_{x_n}(b_n)\right| \geq \epsilon$. We know that $\tilde{a}_\alpha$ is compact so $\exists$ a subsequence $a_{n_i} \to a \in \tilde{a}_\alpha$ which implies that $b_{n_i} \to a$. By assumption $f(a_{n_i},x) \to f(a,x)\ \forall x \in \Omega$, similarly for b_{n_i}. We relabel the subscripts. Then, since $\tilde{f}_\alpha$ is compact and $(f(a_n,x))$, $(f(b_{n_i},x))$ are sequences of functions in this compact function set, $\exists$ a convergent subsequence in $\tilde{f}_\alpha$. This implies $f_{a_n} \to f_a$ and $f_{b_n} \to f_a$ so that $\left\|f_{a_n}\text{-}f_{b_n}\right\|_{\sup} \to 0$. But this is a contradiction since $\left|f_{x_n}(a_n)\text{-}f_{x_n}(b_n)\right| \geq \epsilon \Rightarrow \left|f_{a_n}(x_n)\text{-}f_{b_n}(x_n)\right| \geq \epsilon$.

(Boundedness) This follows by definition of a fuzzy function.

$\Leftarrow$ (Normality) $\tilde{a}$ normal $\Rightarrow \exists a' \in \tilde{a}_1 \Rightarrow \mu_{f(\tilde{a},x)}(f(a',x))=1$.

(Compactness) Let $\alpha \in [0,1]$ and (f_{a_i}) be a sequence of functions in $f(\tilde{a},x)_\alpha$. Then (a_i) is a sequence in $\tilde{a}_\alpha$ which is compact so $\exists(a_{i_n})$ such that $a_{i_n} \to a \in \tilde{a}_\alpha$ and $f_a \in f(\tilde{a},x)_\alpha$. Let $\epsilon > 0$, by definition of equicontinuity $\exists\delta > 0$ such that $|a_{i_n}\text{-}a| < \delta \Rightarrow \left|f_x(a_{i_n})\text{-}f_x(a)\right| < \epsilon \forall x \in \Omega$. Choose N such that $\forall n \geq N, |a_{i_n}\text{-}a| < \delta \Rightarrow \left|f_x(a_{i_n})\text{-}f_x(a)\right| = \left|f(a_{i_n},x)\text{-}f(a,x)\right| < \epsilon \forall x \in \Omega \Rightarrow f_{a_{i_n}} \to f_a$.

(Path Connectedness) Let f_a and $f_b \in f(\tilde{a},x)_\alpha$. Then $a,b \in \tilde{a}$ which is a compact and convex so $\exists$ a path from $F:[0,1] \to R^n$ with $F(0)=a$ and $F(1)=b$. But this also defines a path from f_a to f_b. Let $\epsilon > 0$, then from equicontinuity $\exists \delta_1 > 0$ such that $\forall c, d \in \tilde{a}$, $|c-d| < \delta_1 \Rightarrow |f_x(c)-f_x(d)| < \epsilon \forall x \in \Omega$. By continuity of F $\exists \delta_2 > 0$ such that $\forall x, y \in [0,1]$, $|x-y| < \delta_2 \Rightarrow |F(x)-F(y)| < \delta_1 \Rightarrow |f_x(c)-f_x(d)| < \epsilon \forall x \in \Omega \Rightarrow \|f_c-f_d\|_{\sup} < \epsilon$.

(Pointwise Convexity) This follows immediately since each $\tilde{a}_\alpha$ is closed and connected and f_x is continuous$\Rightarrow \{f_x(a) \mid a \in \tilde{a}_\alpha\}$ is closed and connected in R, therefore convex.□

Example 19 Consider the fuzzy set of bounded functions over R given by $\tilde{f}(\tilde{a},x)=\tilde{a}^2/(\tilde{a}^2+(1-x\tilde{a})^2)$ where $\tilde{a}_\alpha=[0,1]$ $\forall \alpha \in [0,1]$. Using the notation of the above theorem we note that 1) $f_a \equiv 0$ for a=0 2) $\forall a \in (0,1]$ $f_a(x) \to 0$ as $x \to \infty$ and 3) $f_x(a)=1$ when $x=\frac{1}{a}$. Then $\{f_x : \tilde{a}_\alpha \to R\}$ cannot be equicontinuous since if we choose $a \in (0,1]$ but arbitrarily close to zero and let $x=\frac{1}{a}$, then $\left|f_{\frac{1}{a}}(a)-f_{\frac{1}{a}}(0)\right|=1$. Thus this is not a representation of a fuzzy function. It also fails the definition since there is no path from f_a for $a \in (0,1]$ to f_a for a=0. Note, however, that if x is restricted to a compact subset of R, the set of functions is equicontinuous and there is a path to f_a for a=0 and in this case the representation is a fuzzy function.

Theorem 40 Let $f(\tilde{a},x):\Omega \to \mathcal{E}$ represent a fuzzy subset of functions where $\tilde{a}$ is a vector of fuzzy numbers, Ω is compact, and

$$f(a,x) : \Pi_{i=1,n}[(\tilde{a}_i)^-_\alpha, (\tilde{a}_i)^+_\alpha] \times \Omega \to R$$

is continuous $\forall$x$\in \Omega$ and $\forall$a$\in[(\tilde{a}_i)^-_\alpha, (\tilde{a}_i)^+_\alpha]$, $\forall\alpha \in$ [0,1], then f(ã,x) is a representation of a convex fuzzy function.

Proof:

Apply the previous theorem. Since f(a,x) is a continuous mapping of a compact set, it is bounded. Also, it is uniformly continuous, so $\forall\epsilon > 0 \exists\delta > 0$ such that $\|$(a,x)-(b,y)$\| < \delta \Rightarrow \|$f(a,x)-f(b,y)$\| < \epsilon$. Then $\|$a-b$\| < \delta \Rightarrow \|$(a,x)-(b,x)$\| < \delta \forall x \in \Omega \Rightarrow \|f_X$(a)-$f_X(b)\| < \epsilon$ so $\{f_x$:$\tilde{a}_\alpha \rightarrow$ R $|$ x $\in \Omega\}$ is an equicontinuous family of functions. $\square$

A fuzzy function has been defined and it has been shown to be a mapping of a finite dimensional vector space into the space of fuzzy vectors over a finite dimensional vector space. Fuzzy functions include functions with unknown coefficients provided the they are continuous over the support of the parameter. When the range space is the space of fuzzy numbers, the space can be partitioned into the normed space $(\mathcal{L}, \|\cdot\|_{\mathcal{EA}})$. Cauchy sequences will converge in this normed space provided the supports of the fuzzy numbers are uniformly bounded and for many applications, this normed space is the setting into which problems can be meaningfully cast. The usefulness of casting fuzzy optimization problems in this setting is that algorithms have a convergence theory and computations are tractable.

4.2 The Minimum and Minimizer of a Fuzzy Real-valued Function

The process for utilizing possibility theory in optimizing a function with ill-defined parameters was discussed in the introduction to possibility theory in chapter two. First, each unknown parameter of the function is represented by a fuzzy number. Second, for a given point in the decision space, the image of the function is evaluated producing a fuzzy number. Third, the expected average of the fuzzy number representing the image is calculated, a process called defuzzification. Fourth, the function is optimized by finding the point in the decision space that optimizes the expected average. If the range of the fuzzy function is bounded, an algorithm that produces a Cauchy sequence in the space $(\mathcal{L}, \|\cdot\|_{\mathcal{EA}})$ will converge to a solution.

In this process, it is expected that there is a single function being modeled. However the function is unknown. Thus a distribution of possible alternatives for the actual function is used. Each of these possible alternatives may have an optimal value. Thus the distribution of possible functions defines a distribution of possible optimal values. Likewise the points in the decision spaces that produce these possible optima produce a distribution of possible optimizers. In this section we consider the properties of these two distributions. For clarity, we focus on minimization problems.

Definition 33 Let $\tilde{\mathrm{f}}$:$\Omega \to \mathcal{E}$ be a fuzzy function. The **fuzzy minimum** of $\tilde{\mathrm{f}}$ over Ω, is the fuzzy subset, $\tilde{\mathrm{m}}$, of R with membership function

$$\mu_{\tilde{m}}(m) = \begin{cases} \sup\{\alpha \mid \exists \text{ f} \in \tilde{\mathrm{f}}_\alpha \text{ with m=}\inf_{x\in\Omega} \mathrm{f(x)}\} & \text{if } \exists \text{ at least one such f} \\ 0 & \text{otherwise} \end{cases}$$

Theorem 41 Let $\tilde{\mathrm{f}}$:$\Omega \to \mathcal{E}$ be a fuzzy function. Then $\tilde{\mathrm{m}}$, the fuzzy minimum of $\tilde{\mathrm{f}}$, is a fuzzy number.

Proof:

(Normality) $\tilde{\mathrm{f}}$ normal $\Rightarrow \exists \mathrm{f}' \in \tilde{\mathrm{f}}_1$ so that $\inf_{x\in\Omega}\mathrm{f}'(\mathrm{x}) \in \tilde{\mathrm{m}}_1$.

(Compactness) Let $(m_n) \subset \tilde{m}_\alpha$ for $\alpha \in$ [0,1]. Then for each n, $\exists f_n \in \tilde{f}_\alpha$ with $m_n = \inf_{x\in\Omega}\ f_n(x)$. Since $\tilde{\mathrm{f}}_\alpha$ is compact $\exists(\mathrm{f}_{n_i})$ with $\mathrm{f}_{n_i} \to \mathrm{f} \in \tilde{\mathrm{f}}_\alpha$.
Let m=$\inf_{x\in\Omega}$f(x). Let $\epsilon > 0$. $\exists$N such that $\forall n_i >$N $\|\mathrm{f}_{n_i}\text{-f}\|_{\sup} < \frac{\epsilon}{2}$. Assume that for a particular such n_i, $m_{n_i} \leq m$ (note that the argument that follows still holds if we assume $m \leq m_{n_i}$ if we replace f with f_{n_i} and vice versa were ever they appear). For this $n_i \exists x_{n_i} \in \Omega$ with $f_{n_i}(x_{n_i}) - m_{n_i} < \frac{\epsilon}{2}$. But we also have $|f_{n_i}(x_{n_i}) - f(x_{n_i})| < \frac{\epsilon}{2} \Rightarrow f(x_{n_i}) - m_{n_i} < \epsilon$. So by assumption we have $m_{n_i} \leq m \leq f(x_{n_i}) \Rightarrow |\mathrm{m}_{n_i}\text{-m}| < \epsilon$. Thus $m_{n_i} \to m$.

(Convexity) Let m_1 and $m_2 \in \tilde{m}_\alpha$ for $\alpha \in [0,1]$. Then $\exists f_1, f_2 \in \tilde{f}_\alpha$ with $m_1 = \inf_{x\in\Omega} f_1(x)$ and $m_2 = \inf_{x\in\Omega} f_2(x)$. Since $\tilde{f}_\alpha$ is path connected $\exists G : [0,1] \to \tilde{f}_\alpha$ with G continuous and $G(0) = f_1$ and $G(1) = f_2$. Let $F : [0,1] \to R$ be given by $F(r) = inf_{x\in\Omega} G(r)(x)$. Let $s \in [0,1]$ and $\dot{\epsilon} > 0$, G

continuous $\Rightarrow \exists\delta > 0$ such that $r \in [0,1]$ and $|r-s| < \delta \Rightarrow \|G(r) - G(s)\| < \frac{\epsilon}{2}$. Then by application of the same argument as in the compactness proof above (replacing f_{n_i} with $G(r)$ and f with $G(s)$) we have $|F(r) - F(s)| < \epsilon$. Thus F is continuous on $[0,1]$, so $\tilde{m}_\alpha$ is path connected and, thus, convex since $\tilde{m}_\alpha \subset R$. □

Definition 34 Let $\tilde{f}:\Omega \to \mathcal{E}$, where $\tilde{f}$ is a fuzzy function. Then the **fuzzy minimizer** of $\tilde{f}$ over Ω is the fuzzy subset $\tilde{S}$, of Ω, with membership function

$$\mu_{\tilde{S}}(s) = \begin{cases} \sup\{\alpha \mid \exists \text{ f} \in \tilde{\text{f}}_\alpha \text{ with f(s)=} \inf_{x\in\Omega} \text{f(x)}\} & \text{if } \exists \text{ at least one such f} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 42 Let $\tilde{f}:\Omega \to \mathcal{E}$, where $\tilde{f}$ is a fuzzy convex, fuzzy continuous function on Ω and Ω is a compact, convex subset of R^n. Let $\tilde{S}$ be the fuzzy minimizer of $\tilde{f}$ over Ω. Then $\tilde{S}_\alpha$ is connected $\forall\alpha \in$ [0,1].

Proof:

Let s_0 and $s_1 \in \tilde{S}_\alpha$. Then $\exists$ f_0 and f_1 $\in \tilde{f}_\alpha$ with $f_0(s_0)=\inf_{x\in\Omega} f_0(x)=m_0$ and $f_1(s_1)=\inf_{x\in\Omega} f_1(x)=m_1$. Since $\tilde{f}$ is a fuzzy function, $\tilde{f}_\alpha$ is path connected so $\exists$ a path $f_0 \to f_1$, let f_β denote an element along this path. In the proof of Theorem 11 we showed that this path provides a path $m_0 \to m_1$ where $m_\beta=\inf_{x\in\Omega} f_\beta(x)$.

Let $?_\beta=\left\{s \in \Omega \mid f_\beta(s)=m_\beta\right\}$ i.e. it is the set of minimizers of f_β over Ω. Since we assumed f_β is a convex function and Ω convex, we know that $?_\beta$ is convex (see Luenberger [30] Theorem 1 on page 181) and thus connected (convexity$\Rightarrow$path connectedness$\Rightarrow$connectedness, see Munkrees [33] page 155). We also know that $?_\beta$ is non-empty since each f_β is continuous and bounded on Ω by assumption and Ω is compact. Finally, we know $?_\beta$ is closed. To see this let s be a limit point of $?_\beta$. Then s$\in \Omega$ since $?_\beta \subset \Omega$ and Ω is compact (thus closed). Let $(s_i) \subset ?_\beta$ and $s_i \to$s. Then f_β continuous$\Rightarrow f_\beta(s_i)=m_\beta \to f_\beta(s)$, so $f_\beta(s)=m_\beta$ and s$\in ?_\beta$.

We will show that $D(?_\gamma,?_\beta)\to 0$ as $\gamma \to \beta$ where $\gamma, \beta \in$[0,1] and where for two compact sets A,B we define D(A,B) =min$\{\|a\text{-}b\| \mid a \in A, b \in B\}$. The minimum can be used here since A and B are compact. Any sequence of points $(a_i)\subset$A and $(b_i)\subset$B with $\|a_i\text{-}b_i\| \to \inf\{\|a\text{-}b\| \mid a \in A, b \in B\}$ implies $\exists$ a subsequence converging to a point (a,b)$\in$AxB with $\|a\text{-}b\|=\inf\{\|a\text{-}b\| \mid a \in A, b \in B\}$. We will use this to show that the union of our collection of compact connected sets is connected.

Let $\epsilon > 0$ and define T=$\{x \in \Omega \mid D(x,?_\beta)=\epsilon\}$. We show that T is compact. We can choose ϵ small enough so that T is non-empty unless $\tilde{S}_\alpha=\Omega$, in which case we are done since Ω is convex and thus connected. We now show that T is closed. Let w be a limit point of T and $(w_i)\subset$T with $w_i \to$w. Then for each w_i $\exists m_i \in ?_\beta$ with $D(w_i,m_i)=\epsilon$. $?_\beta$ compact implies $\exists$ a subsequence (m_{i_n}) with $m_{i_n} \to$m$\in ?_\beta$. Then $\|w_{i_n}\text{-}m_{i_n}\| \to \|w\text{-}m\|$ so $\|w\text{-}m\|=\epsilon$. Thus $D(w,?_\beta)\leq \epsilon$. Now assume $\exists m_t \in ?_\beta$ with $\|w\text{-}m_t\| < \epsilon$. Since $w_i \to$w and by continuity of the Euclidean norm $\exists w_N$ with $\|w_N - m_t\| < \epsilon$. But this contradicts the fact that $w_N \in T$. Hence $w \in T$ so T is closed. Now T a closed subset of compact set $\Omega \Rightarrow$T compact$\Rightarrow \exists t \in T$ with $f_\beta(t) - m_\beta = \min\left\{f_\beta(x) - m_\beta \mid x \in T\right\}$ since f_β is continuous. Let $\delta = f_\beta(t) - m_\beta$ then $\delta > 0$ since $t \notin ?_\beta$.

We claim that $f_\beta(y) - m_\beta \geq \delta \forall y$ with $D(y, ?_\beta) > \epsilon$. To see this let $z \in ?_\beta$. Since $D(z, ?_\beta) = 0$ and $D(y, ?_\beta) > \epsilon$ and Ω is convex, $\exists \lambda \in [0,1]$ with $\lambda z + (1-\lambda) y \in T$. This follows from the convexity of Ω and continuity of the Euclidean norm. Then $f_\beta(t) \leq f_\beta(\lambda z + (1-\lambda)y)$ by choice of t and $f_\beta(\lambda z + (1-\lambda)y) \leq \lambda f_\beta(z) + (1-\lambda) f_\beta(y)$ since f_β is convex, so $f_\beta(y) - m_\beta \geq f_\beta(t) - m_\beta = \delta$.

We now show that in a neighborhood of β, $D(?_\alpha, ?_\beta) \leq \epsilon$. Since $f_\alpha \to f_\beta$ and $m_\alpha \to m_\beta$ as $\alpha \to \beta$ we can choose a neighborhood of β such that $\forall \alpha \in [0,1]$ in this neighborhood, both $\|f_\alpha - f_\beta\|_{\text{sup}} < \frac{\delta}{2}$ and $|m_\alpha - m_\beta| < \frac{\delta}{2}$. In this neighborhood, assume $\exists r \in ?_\alpha$ and $D(r, ?_\beta) > \epsilon$. Then $\left|f_\beta(r) - m_\beta\right| \geq \delta$ but since $f_\alpha(r) = m_\alpha \Rightarrow |m_\alpha - f_\beta(r)| < \frac{\delta}{2}$ then it must hold that $|m_\alpha - m_\beta| \geq \frac{\delta}{2}$ which is a contradiction. Thus $D(?_\alpha, ?_\beta) \leq \epsilon$.

We have thus shown that $\cup_{\alpha \in [0,1]} ?_\alpha$ is a connected subset of $\tilde{S}_\alpha$ containing two arbitrary points in $\tilde{S}_\alpha$. So $\tilde{S}_\alpha$ must be connected.$\square$

4.3 Method of Minimum Regrets

One possible objective of the fuzzy optimization problem to be to optimize the expected average of the possibility distribution representing the image of the function. In other words, given a fuzzy function $\widetilde{f}(x)$ we define our objective to be $\min EA(\widetilde{f}(x))$. This method is simple to implement. An alternative to this method is the method of minimum regrets. The idea behind this method is to compare the result of the action taken by the decision maker to the optimal result. Of course the optimal result is not known at the time a decision is made, therefore, we define the objective to be to minimize the expected difference between these two quantities.

Consider the problem of minimizing a fuzzy function. The objective is to decide upon a choice of a crisp x in Ω which provides the best possible outcome. If $\tilde{S}$ is the fuzzy minimizer for the fuzzy function then any choice of $x \in \tilde{S}_1$ will have possibility 1 of being the minimizer but, unless $\tilde{S}$ is a crisp vector in X, there is also a possibility that s will not be a minimizer. A tool for implementating the method of minimum regrets is the following:

Definition 35 Let $\tilde{f}:\Omega \to \mathcal{E}$ be a fuzzy function. Define the **maximum possible error** emax:$\Omega\times[0,1]\to$R as follows:

$$\mathrm{emax}(x,\alpha) = \sup\left\{ f(x) - m_f | f \in \widetilde{f}_\alpha \ \text{ and } \ m_f = \mathrm{inf}_{y\in\Omega} f(y)\right\}.$$

Definition 36 Let $\tilde{f}:\Omega \to \mathcal{E}$ be a fuzzy function. Define the **minimum possible error** emin:$\Omega\times$[0,1]$\to$R as follows:

$$\text{emin}(x,\alpha) = \inf\left\{f(x) - m_f | f \in \tilde{f}_\alpha \text{ and } m_f = \inf_{y\in\Omega} f(y)\right\}.$$

These functions are well-defined since each $f\in\tilde{f}_\alpha$ is bounded on Ω and $\tilde{f}_\alpha$ is compact under the supremum norm by definition of a fuzzy function. Thus the supremum is finite since it is taken over a bounded collection of bounded sets.

The maximum possible error gives the maximum possible distance between the function evaluated at x and its minimum value over all functions that are at least α possible. It is natural to want to choose x so as to minimize the maximum possible error. If $\tilde{S}_\alpha$ is nonempty, minimizing at the α-level of possibility can be done by choosing $s\in\tilde{S}_\alpha$ so that $\exists f\in\tilde{f}_\alpha$ with s a minimizer of f and f(s) at the midpoint of $\tilde{m}_\alpha$. The maximum error over all α-level possible functions is then $(m_\alpha^+ - m_\alpha^-)/2$. Minimizing over different α-levels may lead to different solutions. It may be desirable to minimize the maximum possible error for any possible f (α=0). These concepts are illustrated in the following examples.

Example 20 Consider the fuzzy function defined by $\tilde{f}(x)=(\tilde{a}-x)^2+\tilde{a}$

where $\tilde{a}$=[0,1], i.e. it is the unit interval for all α-levels, and Ω=[0,1]. Let

$f_a(x)=(x-a)^2-a$ so $\tilde{f}_0=\{f_a(x) \mid a \in [0,1]\}$. Then

$$\inf_{x\in[0,1]} f_a(x) = a$$

and $\mathrm{emax}(x,0) = \sup_{a\in[0,1]} \left\{(x-a)^2+a-a\right\} = max(x^2,(x-1)^2)$.

On the other hand $\mathrm{emin}(x,0)=0\ \forall x\in[0,1]$ since $\inf_{a\in[0,1]} \left\{(x-a)^2+a-a\right\}=0$. Note that the emax function is minimized at x=.5 where the maximum possible error is .25. For this fuzzy function, this is the best choice if the objective is to minimize the maximum possible error.

Example 21 Consider the fuzzy function defined by $\tilde{f}(x) = (\tilde{a} - x)^2 + \tilde{b}$ where a and b are trapezoidal fuzzy numbers represented by (1.8,2,2.2,2.25) and (.8,1,1.2,2), respectively. For this fuzzy function $\tilde{m}=\tilde{b}$ and $\tilde{S}=\tilde{a}$. For example $\tilde{m}_1=[1,1.2]$ and $\tilde{S}_1=[2,2.2]$ since

$$\tilde{f}_1 = \left\{(a-x)^2+b \mid\ a \in [2,2.2] \text{ and } b \in [1,1.2]\right\}.$$

Thus $f(x) = (2 - x)^2 + 1.2$ is such a function with possibility level 1. Its minimum value is 1.2 and it is minimized at $x = 2$. For $x = 2$ and considering all functions at $\alpha = 1$ possibility level, the maximum possible error occurs if $a = 2.2$ so $f(2) = (2.2 - 2)^2 + b$. The value of b is immaterial. To see this say that $f(x) = (2.2 - x)^2 + 1.1$ which is minimized at $x = 2.2$. Then the error for a choice of $x = 2$ is $f(2) - f(2.2) = 1.14 - 1.1 = .04$. The following table shows the distribution of the values of the maximum possible error function for each choice of x from 2 to 2.2 in steps of .02 and for various possibility levels.

	α					
x	1	.8	.6	.4	.2	0
2.00	.040	.044	.048	.053	.058	.063
2.02	.032	.036	.040	.044	.048	.053
2.04	.026	.029	.032	.036	.040	.058
2.06	.020	.023	.026	.032	.048	.068
2.08	.014	.017	.026	.040	.058	.078
2.10	.010	.020	.032	.048	.068	.090
2.12	.014	.026	.040	.058	.078	.102
2.14	.020	.032	.048	.068	.090	.116
2.16	.026	.040	.058	.078	.102	.130
2.18	.032	.048	.068	.090	.116	.144
2.20	.040	.058	.078	.102	.130	.160

Selecting from this table, a choice of $x = 2.1$ results in the maximum possible

error (.010) being minimized over all level 1 possible functions. But a choice of $x = 2.02$ results in the maximum possible error (.053) being minimized over all possible functions. Which choice is most appropriate would seem to be determined by the context of the problem. In risk management, the later solution might be preferable or a solution for a low α level. In fact, in this context, a problem solution might lie outside the $\tilde{S}_1$ solution set (recall that $\tilde{S}_1 = [2, 2.2]$ in the above example.)

Example 22 Consider the fuzzy function $\tilde{f}(x, y) = (x - \tilde{a})^2 + (y - \tilde{a}^2)^2$ over $[0, 1] \times [0, 1]$ where $\tilde{a}$ is the triangular fuzzy number (.2,.5,.8). For this function $\tilde{m} = 0$, a crisp number, since any possible function of the form $f(x, y) = (x - a)^2 + (y - a^2)^2$ has minimum value 0 at $(x, y) = (a, a^2)$. So $\tilde{S}_\alpha = \{(x, x^2) \mid x \in \tilde{a}_\alpha\}$. For example, $\tilde{S}_0 = \{(x, x^2) \mid x \in [.2, .8]\}$. We see by this example that although $\tilde{S}_\alpha$ is connected it need not be convex. For $\alpha = 0$, the error function is minimized at (.5,.34) where emax((.5,.34),0)=.18. This is because $.5 = (.2 + .8)/2$ so $x = .5$ minimizes $\max\{(x - .2)^2, (x - .8)^2\}$ and $.34 = (.2^2 + .8^2)/2$ so $y = .34$ minimizes $\max\{(y - .2^2)^2, (y - .8^2)^2\}$. Then (.5,.34) minimizes $\max\{(x - a)^2 + (y - a^2)^2 \mid x \in [.2, .8]\}$. Notice that (.5,.34) is not an element of $\tilde{\mathrm{S}}$ so this point minimizes emax but is not a minimizer of any of the possible functions. The point that lies in $\tilde{\mathrm{S}}$ and minimizes emax is $(.544, .544^2)$ where $\mathrm{emax}((.544, .544^2), 0) = .183916$. The possibility that $(.544, .544^2)$ is a minimizer of our fuzzy function is .8533 (since the possibility that $a = .544$ is .8533).

Definition 37 The **possible error** for a given x, $\mathbf{\tilde{e}(x)}$, is the fuzzy set with α-level defined as $\tilde{\mathrm{e}}_\alpha(\mathrm{x})$=[emin(x,$\alpha$),emax(x,$\alpha$)] .

Theorem 43 Let $\tilde{\text{f}}$:$\Omega \to \mathcal{E}$ be a fuzzy function. Then the possible error, ẽ(x), is a fuzzy number.

Proof:

(Normality) Since $\exists f \in \tilde{f}_1$ then $|f(x) - m| \in \tilde{e}(x)_1$ where $m = \inf_{x \in \Omega} f(x)$.

(Compactness) Let $(e_i) \subset \tilde{e}(x)_\alpha$. For each $e_i \exists f_i \in \tilde{f}_\alpha$ with $e_i = |f_i(x) - m_i| \in \tilde{e}(x)_\alpha$ where $m_i = \inf_{x \in \Omega} f_i(x)$. Thus $\exists$ a subsequence $f_{i_n} \to f \in \tilde{f}_\alpha$. From Theorem 41 $m_{i_n} \to m$ so $|f_i(x) - m_i| \to |f(x) - m| \in \tilde{e}(x)_\alpha$.

(Convexity) Let $e_1, e_2 \in \tilde{e}(x)_\alpha$, then $\exists f_1, f_2 \in \tilde{f}_\alpha$ with $e_i = |f_i(x) - m_i|$ where $m_i = \inf_{x \in \Omega} f_i(x)$. $\exists$ a path $G : [0,1] \to \tilde{f}_\alpha$ mapping f_1 to f_2. Then $G_x : [0,1] \to R$ defined by $G_x(a) = G(a)(x)$ is continuous as is the function $F : [0,1] \to R$ from Theorem 41 (mapping to the infimum of $G(a)$). Then $H : [0,1] \to R$ defined by $H(a) = |G_x(a) - F(a)|$ is continuous with $H(0) = e_1$ and $H(1) = e_2$ so $\forall \lambda \in (0,1) \exists r \in [0,1]$ with $H(r) = \lambda e_1 + (1 - \lambda) e_2$. $\Box$

We see that the possible error is a possibility distribution for the error that may occur for a given choice of x as the function input value. An objective in choosing x may be to achieve the most favorable possible error distribution. To make such a choice one needs a method of rating these distributions. One approach is to minimize the expected error as estimated by the EA functional as follows:

$$EA(\tilde{e}(x)) = \frac{1}{2}\int(\tilde{e}_{\alpha}^{-} + \tilde{e}_{\alpha}^{+})d\alpha$$

Using this functional one can then redefine the minimization problem as:

For fuzzy function $\tilde{f}$, find x which minimizes EA($\tilde{e}$(x)).

Example 23 As a final example consider the fuzzy function in Example 19 defined by $\tilde{f}(x) = (\tilde{a} - x)^2 + \tilde{b}$ where a and b are trapezoidal fuzzy numbers represented by (1.8,2,2.2,2.25) and (.8,1,1.2,2) respectively. Now we wish to find the minimum of EA($\tilde{e}$(x)) for this function. This can be handled analytically in a piecewise manner. We will only show the calculations for the interval in R where we know the answer lies. On the interval $2.025 \le x \le 2.1$, for $(.15\alpha + 4.05)/2 > x, \tilde{e}_{\alpha}(x) = [0, (-.05\alpha + 2.25 - x)^2]$ for $(.15\alpha + 4.05)/2 \le x, \tilde{e}_{\alpha}(x) = [0, (.2\alpha + 1.8 - x)^2]$ Then

$$\begin{aligned} &EA(\tilde{e}(x)) \\ &\quad = \frac{1}{2}\left(\int_{(2x-4.05)/.15}^{1}(-.05\alpha + 2.25 - x)^2 d\alpha + \int_{0}^{(2x-4.05)/.15}(.2\alpha + 1.8 - x)^2 d\alpha\right) \end{aligned}$$

In the interval of evaluation, this function is minimized at x=2.057 where EA($\tilde{e}$(x))=1.4573×10^{-2}. We consider this x to be the solution using the method of minimum regrets.

For comparison we calculate the minimum using the expected average as follows:

$$\tilde{f}(x)^+_\alpha = \max\left\{\left(\tilde{a}^+_\alpha - x\right)^2, \left(\tilde{a}^-_\alpha - x\right)^2\right\} + \tilde{b}^+_\alpha =$$

$$\max\left\{(2.25 - .05\alpha - x)^2, (1.8 + .2\alpha - x)^2\right\} + 2 - .8\alpha$$

$$\tilde{f}(x)^-_\alpha = \min\left\{\left(\tilde{a}^+_\alpha - x\right)^2, \left(\tilde{a}^-_\alpha - x\right)^2\right\} + \tilde{b}^-_\alpha =$$

$$\min\left\{(2.25 - .05\alpha - x)^2, (1.8 + .2\alpha - x)^2\right\} + .8 + .2\alpha$$

$$\begin{aligned}
&\tilde{f}(x)^+_\alpha + \tilde{f}(x)^-_\alpha \\
&= (2.25 - .05\alpha - x)^2 + (1.8 + .2\alpha - x)^2 + 2 - .8\alpha + .8 + .2\alpha \\
&= 11.103 - .105\alpha - 8.1x + .0425\alpha^2 - .3\alpha x + 2.0x^2
\end{aligned}$$

$$\begin{aligned}
&EA(\tilde{f}(x)) \\
&= \frac{1}{2}\int_0^1 \left(11.103 - .105\alpha - 8.1x + .0425\alpha^2 - .3\alpha x + 2.0x^2\right) d\alpha \\
&= 5.5323 - 4.125x + x^2
\end{aligned}$$

which is minimized at $x = 2.0625$ where $EA(\tilde{f}(x)) = 1.2784$.

5. Fuzzy Linear Programming using a Penalty Method

The techniques discussed in this thesis are illustrated by examining the linear programming problem. Starting with a standard form of the linear programming problem each constant in the problem is replaced with a fuzzy number. The objective and constraints are reformatted into an unconstrained fuzzy function by penalizing the objective for possible constraint violations. The range of this fuzzy function lies in the space of fuzzy numbers. The objective is then redefined as maximizing the expected average of the image of this fuzzy function. It is shown that this objective defines a concave function which, therefore, can be maximized globally. An algorithm for finding the maximum is presented.

In constrained optimization problems where uncertainty is characterized using possibility distributions, it may be unavoidable and/or advantageous to consider solutions that have a non-zero possibility of violating one or more of the constraints. This can be done by considering the cost of a constraint violation in the problem formulation. In the following discussion such a formulation of the linear programming problem is examined where the constant terms in the problem may not be known precisely. To incorporate this type of uncertainty into the model each constant in the problem is replaced with a fuzzy number. Next the possibility of a constraint violation is incorporated into the model. This is done by replacing each constraint with a term in the objective function that reduces the objective by the cost of the violation. From this an unconstrained fuzzy function optimization problem arises. As we have seen, the image of this fuzzy function is a fuzzy number. Thus the objective is to, in some sense, find the optimal fuzzy number. To do this we choose the EA functional as an estimate of the expected value of the random variable underling the possibilitic number. From this point of view, which is the approach of this section, the objective is to maximize the expected average of the fuzzy number that represents the possible outcomes for a given action. A gradient ascent algorithm for finding the solution to this problem is developed.

5.1 Problem Formulation

The following is the form of the crisp linear programming problem considered herein:

Maximize $c^T x$

Subject to $Ax \leq b$

$x \geq 0$

Where c and $x \in R^n$ and $b \in R^m$.

If there are uncertainties about any of the components of A and/or b the possibility of a constraint violation cannot be avoided unless the problem restricts x to the worst/best possible case (optimistic, pessimistic linear program - see Lodwick [27]). To take into account the possibility of a constraint violation each constraint is replaced with a penalty term in the objective function together with the corresponding uncertainty in the coefficients. The actual penalty term will be problem dependent though its generic representation is developed and analyzed. For this thesis we will treat constraints as resources and assume that if a resource is exceeded it can be replenished at a cost that is linear with respect to the amount of the violation. The incorporation of truly hard constraints is easily handled within our approach. If the resource constraint is "hard" the penalty is ∞. The exception to the above treatment of constraints is that $x \geq 0$ is considered a crisp constraint. In other words, we will replace the following constraint:

$$A_i x \leq b_i$$

by subtracting the following penalty term from the objective function,

$$d_i \max(0, A_i x - b_i)$$

where each $d_i > 0$ is the cost per unit of violation of the right-hand side value. The objective function now takes the following form, where the maximum is taken component wise and d$\in R^m$ and each component is positive:

$$f(x) = c^T x - d^T \max(0, Ax-b)$$

Now we are in a position to replace each component of c and b and each coefficient of A with a fuzzy number to get the following:

$$\tilde{f}(x) = \tilde{c}^T x - \tilde{d}^T \max(0, \tilde{A}x-\tilde{b})$$

where max is handled component wise using the extension principle. If we restrict x to a compact subset R^n, then $\tilde{f}$ defines a fuzzy function. It has been shown that the image at any vector x, $\tilde{f}(x)$, is a fuzzy number. This fuzzy number is completely characterized by it's $\alpha - cuts$:

$$\tilde{f}(x)_\alpha = \left\{ c^T x - d^T \max(0, Ax - b) \mid c, d, A, b \in \tilde{c}_\alpha, \tilde{d}_\alpha, \tilde{A}_\alpha \text{ and } \tilde{b}_\alpha \text{ respectively} \right\}$$

This fuzzy number provides the possibility distribution for the outcome of taking action x. We wish to find the most favorable possibility distribution over all possible actions. As has been discussed, using the expected average of the fuzzy number as the basis of comparing two fuzzy numbers makes sense for a decision maker whose utility for an interval of possible values is the midpoint of the interval. This provides the additional advantage of insuring our algorithm will converge. Our new optimization problem is as follows:

$$\text{maximize } EA(\tilde{f}(x)) = EA(\tilde{c}^T x - \tilde{d}^T \max(0, \tilde{A}x-\tilde{b}))$$

This objective function requires the α-cuts of our fuzzy number $\tilde{\mathrm{f}}(\mathrm{x})$ given the definition of expected average. But this is straightforward given the alpha cuts of our fuzzy coefficients because of the linearity of our original problem and the nonnegativity of x. Therefore,

$\tilde{\mathrm{f}}^{+}_{\alpha}(\mathrm{x}) = (\tilde{\mathrm{c}}^{+}_{\alpha})^{\mathrm{T}}\mathrm{x}$ - $(\tilde{\mathrm{d}}^{-}_{\alpha})^{\mathrm{T}}\max(0,\tilde{\mathrm{A}}^{-}_{\alpha}\mathrm{x}\text{-}(\tilde{\mathrm{b}}^{+}_{\alpha}))$

$\tilde{\mathrm{f}}^{-}_{\alpha}(\mathrm{x}) = (\tilde{\mathrm{c}}^{-}_{\alpha})^{\mathrm{T}}\mathrm{x}$ - $(\tilde{\mathrm{d}}^{+}_{\alpha})^{\mathrm{T}}\max(0,\tilde{\mathrm{A}}^{+}_{\alpha}\mathrm{x}\text{-}(\tilde{\mathrm{b}}^{-}_{\alpha}))$

These two numbers define the right and left end-points of the α-cut of the fuzzy function evaluated at x where $\tilde{\mathrm{f}}^{+}_{\alpha}(\mathrm{x})$ is called the **optimistic** value of $\tilde{f}(x)$ at possibility level α and $\tilde{\mathrm{f}}^{-}_{\alpha}(\mathrm{x})$ is called the **pessimistic** value. The modified problem now becomes:

$$\text{Maximize } \mathrm{EA}(\tilde{\mathrm{f}}(\mathrm{x})) = \tfrac{1}{2}\int_0^1 (\tilde{\mathrm{f}}^{-}_{\alpha}(\mathrm{x}) + \tilde{\mathrm{f}}^{+}_{\alpha}(\mathrm{x}))d\alpha =$$

$$\frac{1}{2}\int_0^1 ((\tilde{c}^{-}_{\alpha})^{\mathrm{T}}x + (\tilde{c}^{+}_{\alpha})^{\mathrm{T}}x - (\tilde{d}^{-}_{\alpha})^{\mathrm{T}}max(0, \tilde{A}^{-}_{\alpha}x - (\tilde{b}^{+}_{\alpha})) - (\tilde{d}^{+}_{\alpha})^{\mathrm{T}}max(0, \tilde{A}^{+}_{\alpha}x - (\tilde{b}^{-}_{\alpha}))d\alpha \tag{5.1}$$

5.2 Properties of the Fuzzy Optimization Problem

Theorem 44 The fuzzy optimization problem as defined by (5.1) is concave.

Proof:

To see that mapping $\mathrm{EA}(\tilde{\mathrm{f}}(\mathrm{x})):\mathrm{R}^{\mathrm{n}} \rightarrow \mathrm{R}$ defined above is concave only the terms involving the maximum operator need be considered. Let $\beta \in (0,1)$, x and y$\in$R^n and A_i a row of matrix A. Due to the properties of the integral all that needs to be shown is that

$$\max(0,A_i^T(\beta x+(1-\beta)y)-b)\leq \beta\max(0,A_i^T x-b)+(1-\beta)\max(0,A_i^T y-b) \quad \cdot$$

Let $A_i^T x=z$ and $A_i^T y=w$. Then the above is equivalent to

$$\max(0,\beta z+(1-\beta)w-b)\leq \beta\max(0,z-b)+(1-\beta)\max(0,w-b)$$

If $z, w > b$ then all terms are greater than zero and equality holds. If $z, w < b$ then all terms are less than zero so the maximum is zero and again equality holds. Assume without loss of generality that $z > b$ and $w < b$. Then

$$\text{L.H.S.} = \max(0,\beta z+(1-\beta)w-b)\leq\max(0,\beta z+(1-\beta)b-b)=\beta(z-b) = \text{R.H.S.} \quad \Box.$$

Since the fuzzy optimization problem as defined by (5.1) is concave, a solution should be obtainable using a gradient ascent technique if the problem is bounded. A test for boundedness of (5.1) is as follows:

Theorem 45 If the union of the feasible sets for all possible crisp formulations of the original linear programming problem is bounded then the fuzzy optimization problem is bounded if and only if for all j= 1 to n

$$EA(\tilde{c}_j) \leq \frac{1}{2}\int_0^1 \left[(\tilde{d}_\alpha^-)^T(\tilde{A}^j)_\alpha^- + (\tilde{d}_\alpha^+)^T(\tilde{A}^j)_\alpha^+\right] d\alpha = EA(\tilde{d}^T\tilde{A}^j) \tag{5.2}$$

Proof:

If the union of all possible formulations are bounded then we can focus on the objective function for those x that are outside of this union. For this region of R^n each maximum in (5.1) is $\widetilde{A}x - \widetilde{b}$ and (5.1) becomes

$$\frac{1}{2}\int_0^1 ((\tilde{c}_\alpha^-)^{\mathrm{T}} x + (\tilde{c}_\alpha^+)^{\mathrm{T}} x - (\tilde{d}_\alpha^-)^{\mathrm{T}} (\tilde{A}_\alpha^- x - (\tilde{b}_\alpha^+)) - (\tilde{d}_\alpha^+)^{\mathrm{T}} (\tilde{A}_\alpha^+ x - (\tilde{b}_\alpha^-)) d\alpha$$

Thus

$$\partial(EA(\tilde{f}(x)))/\partial x_j = \frac{1}{2}\int_0^1 (\tilde{c}_j)_\alpha^- + (\tilde{c}_j)_\alpha^+ - (\tilde{d}_\alpha^-)^{\mathrm{T}} (\tilde{A}^j)_\alpha^- - (\tilde{d}_\alpha^+)^{\mathrm{T}} (\tilde{A}^j)_\alpha^+) d\alpha \quad (5.3)$$

and the condition of the theorem follows by requiring that the gradient be nonpositive.□

This theorem implies that the partial with respect to x_j is negative when all constraints are fully violated. This means that the expected increase in the original objective function for an increase in variable x_j $(EA(\tilde{c}_j))$ must be less than the expected increase in cost of replenishing the resources needed to obtain the increase $(EA(\widetilde{d}^T \widetilde{A}^j))$.

Note that if the conditions of the above theorem are met then we know we have convergence of Cauchy sequences by Theorem 37.

We can find a bound on the solution to the fuzzy optimization problem of (5.1). The absolute (crisp) upper bound on x is given in the next theorem.

Theorem 46 If the fuzzy optimization problem is bounded, then $\forall j = 1, n$

$$x_j \leq \max\left\{(b_i)_0^+/(a_{ij})_0^- \mid (a_{ij})_0^- \neq 0, i = 1, m\right\}$$

Proof:

This is an immediate consequence of the prior theorem since if x_j does not satisfy this condition then the maximum operator in (5.1) disappears and the partial at x_j as given by (5.3) will be negative.$\square$

5.3 An Algorithm

This section provides the details of a line search algorithm for finding the maximum of the fuzzy optimization problem. The first step in the algorithm determines if any of the components of the terms max(0,$\tilde{A}^-_\alpha$x-($\tilde{b}^+_\alpha$)) and max(0,$\tilde{A}^+_\alpha$x-($\tilde{b}^-_\alpha$)) become active for some $\alpha \in (0,1)$. The second step of the algorithm calculates the values of these α's. The third step utilizes the calculated α's to determine the gradient of the objective function. This gradient is then used as the direction in which to search for the optimal value. Note that since the problem is concave, a local optimum is a global optimum.

Step One

The objective of this step is to determine which constraints are violated at some $\alpha - cut$ for a given x. Let the set Ω consist of the indices for the constraints that will be violated for some $\alpha \in (0,1)$ using the pessimistic values of $\widetilde{A}$ and $\tilde{b}$. Let Ψ consist of the indices for the constraints that are violated for all α using the pessimistic values. Let sets ? and Λ be the same tests but using the optimistic values of $\widetilde{A}$ and $\widetilde{b}$. Formally, these sets are defined as follows:

$\Omega=\Big\{ i \mid (\tilde{A}_i)_0^+ \mathrm{x} \geq (\tilde{\mathrm{b}}_i)_0^-) \text{ and } (\tilde{A}_i)_1^+ \mathrm{x}¡(\tilde{\mathrm{b}}_i)_1^-) \Big\}$,

$\Psi=\Big\{ i \mid (\tilde{A}_i)_1^+ \mathrm{x} \geq (\tilde{\mathrm{b}}_i)_1^-) \Big\}$,

? $=\Big\{ i \mid (\tilde{A}_i)_1^- \mathrm{x} \geq (\tilde{\mathrm{b}}_i)_1^+) \text{ and } (\tilde{A}_i)_0^- \mathrm{x}¡(\tilde{\mathrm{b}}_i)_0^+) \Big\}$ and

$\Lambda=\Big\{ i \mid (\tilde{A}_i)_0^- \mathrm{x} \geq (\tilde{\mathrm{b}}_i)_0^+) \Big\}$.

Step Two

For each of the constraints that will be violated for some $\alpha \in (0,1)$ we identify the α at which the constraint is first violated:

For each i$\in \Omega$ let α_i^+ solve $((\tilde{A}_i)^+_{\alpha_i^+})^T \mathrm{x}=(\tilde{\mathrm{b}}_i)^-_{\alpha_i^+})$, and

for each i$\in$?, let α_i^- solve $((\tilde{A}_i)^-_{\alpha_i^-})^T \mathrm{x}=(\tilde{\mathrm{b}}_i)^+_{\alpha_i^-})$.

Step Three

Calculate the gradient of EA($\tilde{\mathrm{f}}$(x)) by the following formula,

$$\partial(\mathrm{EA}(\tilde{\mathrm{f}}(\mathrm{x})))/\partial \mathrm{x}_j=\mathrm{EA}(\tilde{\mathrm{c}}_j)$$

$$-\textstyle\sum_{i\in\Omega} \frac{1}{2} \int_0^{\alpha_j^+} (\tilde{\mathrm{d}}_i)^+_\alpha (\tilde{\mathrm{a}}_{ij})^+_\alpha \mathrm{d}\alpha$$

$$-\textstyle\sum_{i\in\Psi} \frac{1}{2} \int_0^1 (\tilde{\mathrm{d}}_i)^+_\alpha (\tilde{\mathrm{a}}_{ij})^+_\alpha \mathrm{d}\alpha$$

$$-\textstyle\sum_{i\in\Gamma} \frac{1}{2} \int_{\alpha_j^-}^1 (\tilde{\mathrm{d}}_i)^-_\alpha (\tilde{\mathrm{a}}_{ij})^-_\alpha \mathrm{d}\alpha$$

$$-\textstyle\sum_{i\in\Lambda} \frac{1}{2} \int_0^1 (\tilde{\mathrm{d}}_i)^-_\alpha (\tilde{\mathrm{a}}_{ij})^-_\alpha \mathrm{d}\alpha$$

Step Four

Test gradient for sufficiently close to zero.

If sufficiently close, stop otherwise continue.

Step Five

Line search (e.g. see Luenberger [30]) in the direction of the gradient.

Return to step one.

The above algorithm requires special handling of the constraint violation due to the use of the maximum operator in the integral. Alternatively, max(0,x) could be replaced by

$$\frac{\sqrt{x^2+\epsilon^2}+x}{2}$$

where ϵ is a very small constant.

Appendix B provides formulas for implementing the gradient ascent algorithm when the coefficients of the original linear programming problem are replaced by trapezoidal fuzzy numbers.

5.4 Example

The following crisp problem is from Luenberger [30].

Maximize $2x_1+x_2$

Subject to $x_1+\frac{8}{3}x_2 \leq 4$

$x_1+x_2 \leq 2$

$2x_1 \leq 3$

$x_1,x_2 \geq 0$

The solution to the crisp problem is 3.5 at (1.5,.5).

Assume that the penalty for violating the three constraints are 3,2,3 per unit of violation respectively. We replace each term in the problem with the triangular fuzzy number with $\alpha - cut$ as follows:

$\tilde{\mathrm{w}}_\alpha = [w - .5 + .5\alpha, w + .5 - .5\alpha]$.

Thus we replace the number 2 with the fuzzy number $\widetilde{2}$ with $\alpha - cut$ $\widetilde{2}_\alpha = [1.5 + .5\alpha, 2.5 - .5\alpha]$. We interpret this to mean that the probability is α that $\widetilde{2}_\alpha$ contains the range of possible values for $\widetilde{2}$ (as defined in chapter two). For example, the probability is .5 that the range of possible values for $\widetilde{2}$ is a subset of $[1.75, 2.25]$. With these replacements our penalized fuzzy function is:

$\tilde{f}(x) = \tilde{2}x_1 + \tilde{1}x_2$

$-\tilde{3}\max[0, \tilde{1}x_1 + \tilde{\frac{8}{3}}x_2 - \tilde{4}]$

$-\tilde{2}\max[0, \tilde{1}x_1 + \tilde{1}x_2 - \tilde{2}]$

$-\tilde{3}\max[0, \tilde{2}x_1 - \tilde{3}]$.

Our reformulated problem becomes:

Maximize $\frac{1}{2}\int_0^1 \left[\tilde{f}_\alpha^+(x) + \tilde{f}_\alpha^-(x)\right] d\alpha$ where

$\tilde{f}_\alpha^+(x) = (2.5 - .5\alpha)x_1 + (1.5 - .5\alpha)x_2$

$-(2.5 + .5\alpha)\max[0, (.5 + .5\alpha)x_1 + (\frac{8}{3} - .5 + .5\alpha)x_2 - (4.5 - .5\alpha)]$

$-(1.5 + .5\alpha)\max[0, (.5 + .5\alpha)x_1 + (.5 + .5\alpha)x_2 - (2.5 - .5\alpha)]$

$-(2.5+.5\alpha)\max[0,(1.5+.5\alpha)x_1-(3.5-.5\alpha)]$

and

$\tilde{f}^-_\alpha(x)=(1.5+.5\alpha)x_1+(.5+.5\alpha)x_2$

$-(3.5-.5\alpha)\max[0,(1.5-.5\alpha)x_1+(8/3+.5-.5\alpha)x_2-(3.5+.5\alpha)]$

$-(2.5-.5\alpha)\max[0,(1.5-.5\alpha)x_1+(1.5-.5\alpha)x_2-(1.5+.5\alpha)]$

$-(3.5-.5\alpha)\max[0,(2.5-.5\alpha)x_1-(2.5+.5\alpha)]$.

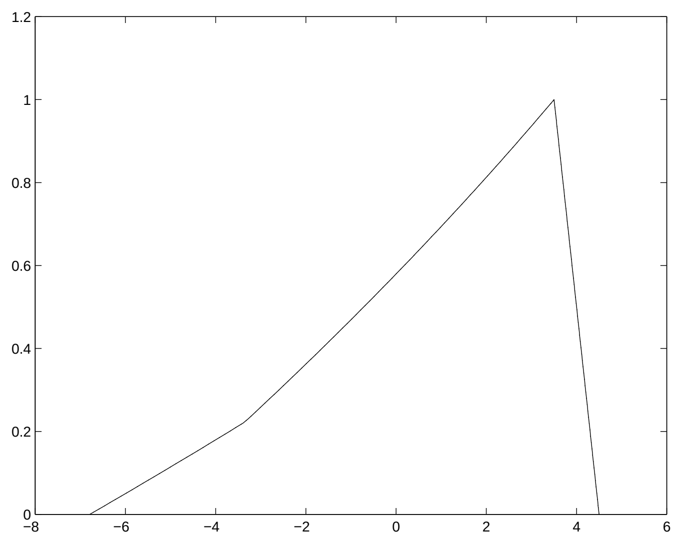

Figure 5.1. Image of fuzzy function at $(1.5, 0.5)$.

The image of this fuzzy function at (1.5,.5), a fuzzy number, is shown in Figure 5.1. The expected average of this fuzzy number, $\mathrm{EA}(\widetilde{f}(1.5,.5))$, is 1.5192. This is not the optimal value of the fuzzy optimization problem. The optimal value of the fuzzy optimization problem is found at (1.1,.4372) where

EA($\widetilde{f}(1.1, .4372)$) =2.2794 (see Appendix A for the details on the implementation of the ascent algorithm for this problem).

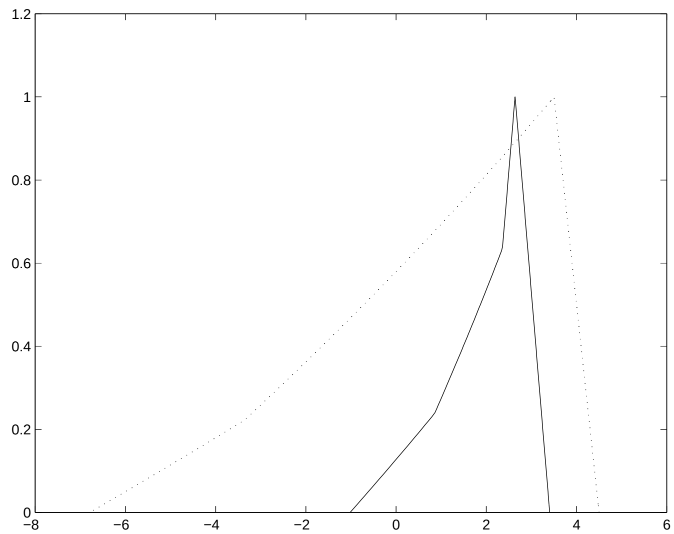

Figure 5.2. Image of fuzzy function at $(1.5, 0.5)$ and at optmial point $(1.1, 0.4372)$.

Figure 5.2 shows the fuzzy number $\widetilde{f}(1.5, .5)$ and the fuzzy number $\widetilde{f}(1.1, .4372)$ which is optimal with respect to the EA functional. From Figure 5.2 you can see the trade-off that is evident in this formulation of fuzzy programing. Recall that the optimistic values of our fuzzy function are the right hand sides of each α-level of the fuzzy number image and the pessimistic values are the left hand sides. The optimistic values of $\widetilde{f}(1.1, .4372)$ are less than the optimistic values of $\widetilde{f}(1.5, .5)$ but this is more than offset by increases in the pessimistic values. Recalling that possibility levels are upper bounds on probabilities one can also state that there is at most a 10% probability that a negative result is possible at (1.1,.4372) compared to at most 60% probability for (1.5,.5).

6. Conclusion

Fuzzy set theory has been an area of study for some time now but it is still controversial. The principle source of the controversy is the lack definiteness in assigning membership values. Even so, much has been done with the theory stemming from the simplicity of the rules of combination and the ability of the theory to model linguistic data. Possibility theory is a subset of fuzzy set theory and inherits from it the rules of combination. This makes it an attractive candidate for modeling uncertainty.

In this thesis we have examined the application of possibility theory to optimization problems involving uncertainty. Of particular interest is the modeling of parameterized functions where the parameters are not known but can be represented by probability distributions. The value of such a function is it self a random variable and it is in this random variable we are ultimately interested. We showed how to gain information about this random variable by constructing a possibility distribution for the variable.

We began this thesis by addressing the concern many researchers have expressed for possibility theory by constructing membership values directly from probability theory. We saw that a possibility distribution constructed in this way gave a partial representation for the random variable representing the value of the uncertain function being optimized. In this setting, there is no subjectivity in assigning membership values. They are a direct consequence of the probability distributions chosen to represent the parameters of the function. However, there is choice in constructing possibilty distributions this way. There are many, perhaps an infinite number of different possibility distributions, that can be constructed to represent a given random variable. Therefore, more research needs to be done regarding how the possibility distributions are constructed. Is there a single method that gives the best result? Do the set of all possibility distributions for a random variable completely characterize the probability distribution? If so, is there a sequence of possibility distributions that converge to a complete charaterization of the probability distribution?

Additional areas of research include further analysis of the EE functional as an expected value estimator, examination of variance estimators, determining if there is a method for constructing the possibility distributions for the unknown parameters that results in the EE or EA functionals producing better estimates of the expected value, examination of expected utility functionals (where a decision makers utility for an interval of possible values is used as a decision criteria instead of the midpoint), study of constrained nonlinear optimization techniques using the EA or EE functionals, further study of the method of minimum regrets and finally, application of the methods to real world problems and a comparison of the results to other known methods.

A. APPENDIX Details of Fuzzy Linear Programming Example

The test for boundedness and the details of the application of the gradient ascent algorithm to our example is presented here. We will need the following for our algorithm:

$$A_0^- = \begin{bmatrix} .5 & 2.1667 \\ .5 & .5 \\ 1.5 & 0 \end{bmatrix} \quad A_1^{+\text{ or }-} = \begin{bmatrix} 1 & 2.6667 \\ 1 & 1 \\ 2 & 0 \end{bmatrix} \quad A_0^+ = \begin{bmatrix} 1.5 & 3.1167 \\ 1.5 & 1.5 \\ 2.5 & 0 \end{bmatrix}$$

$$b_0^- = \begin{bmatrix} 3.5 \\ 1.5 \\ 2.5 \end{bmatrix} \quad b_1^{+\text{ or }-} = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} \quad b_0^+ = \begin{bmatrix} 4.5 \\ 2.5 \\ 3.5 \end{bmatrix}$$

$$d_0^- = \begin{bmatrix} 2.5 \\ 1.5 \\ 2.5 \end{bmatrix} \quad d_1^{+\text{ or }-} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} \quad d_0^+ = \begin{bmatrix} 3.5 \\ 2.5 \\ 3.5 \end{bmatrix}$$

Test for Boundedness

We first test to see if our problem is bounded:

For i=1,

EA($\tilde{c}_1$) = 2 and

$\frac{1}{2}\int_0^1 \left[(\tilde{d}_\alpha^-)^T(\widetilde{A}^1)_\alpha^- + (\tilde{d}_\alpha^+)^T(\widetilde{A}^1)_\alpha^+\right] d\alpha =$

$$\frac{1}{2}\int_0^1 \left(\begin{array}{c} \begin{pmatrix} 2.5+.5\alpha & 1.5+.5\alpha & 2.5+.5\alpha \end{pmatrix} \begin{pmatrix} .5+.5\alpha \\ .5+.5\alpha \\ 1.5+.5\alpha \end{pmatrix} + \\ \begin{pmatrix} 3.5-.5\alpha & 2.5-.5\alpha & 3.5-.5\alpha \end{pmatrix} \begin{pmatrix} 1.5-.5\alpha \\ 1.5-.5\alpha \\ 2.5-.5\alpha \end{pmatrix} \end{array} \right) d\alpha =$$

$\frac{1}{2}\int_0^1 (23.5 - 3.0\alpha + 1.5\alpha^2) d\alpha = 11.25$

Thus EA($\tilde{c}_1$) $< \frac{1}{2}\int_0^1 \left[(\tilde{d}_\alpha^-)^T(\widetilde{A}^1)_\alpha^- + (\tilde{d}_\alpha^+)^T(\widetilde{A}^1)_\alpha^+\right] d\alpha.$

For i=2,

EA($\tilde{c}_2$) = 1 and

$\frac{1}{2}\int_0^1 \left[(\tilde{d}_\alpha^-)^T(\widetilde{A}^2)_\alpha^- + (\tilde{d}_\alpha^+)^T(\widetilde{A}^2)_\alpha^+\right] d\alpha =$

$$\frac{1}{2}\int_0^1 \left(\begin{array}{c} \begin{pmatrix} 2.5+.5\alpha & 1.5+.5\alpha & 2.5+.5\alpha \end{pmatrix} \begin{pmatrix} 2.1667+.5\alpha \\ .5+.5\alpha \\ 0 \end{pmatrix} + \\ \begin{pmatrix} 3.5-.5\alpha & 2.5-.5\alpha & 3.5-.5\alpha \end{pmatrix} \begin{pmatrix} 2.6667-.5\alpha \\ 1.5-.5\alpha \\ 0 \end{pmatrix} \end{array} \right) d\alpha =$$

$\frac{1}{2}\int_0^1 (19.25 - 1.75\alpha + \alpha^2) d\alpha = 9.3542$

Thus our problem is bounded.

An Upper Bound on x_j

We compute an upper bound on x as follows:

$x_1 \leq \max\{4.5/.5, 2.5/.5, 3.5/1.5\} = 9.0$

and

$x_2 \leq \max\{4.5/2.1667, 2.5/.5\} = 5.0$

Application of the Ascent Algorithm

Let us start at x=(1.5,.5), where EA($\widetilde{f}$(x))=

$$\int_0^1 \tfrac{1}{2}((2.5-.5\alpha)1.5+(1.5-.5\alpha).5)d\alpha$$

$$-\int_0^1 \tfrac{1}{2}(2.5+.5\alpha)\max[0,(.5+.5\alpha)1.5+(\tfrac{8}{3}-.5+.5\alpha).5-(4.5-.5\alpha)]d\alpha$$

$$-\int_0^1 \tfrac{1}{2}(1.5+.5\alpha)\max[0,(.5+.5\alpha)1.5+(.5+.5\alpha).5-(2.5-.5\alpha)]d\alpha$$

$$-\int_0^1 \tfrac{1}{2}(2.5+.5\alpha)\max[0,(1.5+.5\alpha)1.5-(3.5-.5\alpha)]d\alpha$$

$$+\int_0^1 \tfrac{1}{2}((1.5+.5\alpha)1.5+(.5+.5\alpha).5)d\alpha$$

$$-\int_0^1 \tfrac{1}{2}(3.5-.5\alpha)\max[0,(1.5-.5\alpha)1.5+(\tfrac{8}{3}+.5-.5\alpha).5-(3.5+.5\alpha)]d\alpha$$

$$-\int_0^1 \tfrac{1}{2}(2.5-.5\alpha)\max[0,(1.5-.5\alpha)1.5+(1.5-.5\alpha).5-(1.5+.5\alpha)]d\alpha$$

$$-\int_0^1 \tfrac{1}{2}(3.5-.5\alpha)\max[0,(2.5-.5\alpha)1.5-(2.5+.5\alpha)]d\alpha = 1.5192$$

We will look for an improvement in this value by using the gradient ascent algorithm as follows:

Step One:

Determine Ω

$$\mathrm{A}_0^+ x = \begin{bmatrix} 3.8334 \\ 3 \\ 3.75 \end{bmatrix} \geq^{\Gamma} b_0^- = \begin{bmatrix} 3.5 \\ 1.5 \\ 2.5 \end{bmatrix} \text{ and}$$

$$\mathrm{A}_1^+ x = \begin{bmatrix} 2.8334 \\ 2 \\ 3 \end{bmatrix} <^{\Gamma} b_1^- = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} \begin{matrix} yes \\ no \\ no \end{matrix}$$

$\Omega = \{1\}$

Determine Ψ

$$\mathrm{A}_1^+ x = \begin{bmatrix} 2.8334 \\ 2 \\ 3 \end{bmatrix} \geq^{\Gamma} b_1^- = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} \begin{matrix} no \\ yes \\ yes \end{matrix}$$

$\Psi = \{2, 3\}$

Determine ?

$$\mathrm{A}_1^- x = \begin{bmatrix} 2.8334 \\ 2 \\ 3 \end{bmatrix} \geq^{\Gamma} b_1^+ = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} \text{ and}$$

$$\mathrm{A}_0^- x = \begin{bmatrix} 1.8334 \\ 1 \\ 2.25 \end{bmatrix} <^{\Gamma} b_0^+ = \begin{bmatrix} 4.5 \\ 2.5 \\ 3.5 \end{bmatrix} \begin{matrix} no \\ yes \\ yes \end{matrix}$$

? $= \{2, 3\}$

Determine Λ

$$\mathrm{A}_0^- x = \begin{bmatrix} 1.8334 \\ 1 \\ 2.25 \end{bmatrix} \geq^\Gamma b_0^+ = \begin{bmatrix} 4.5 \\ 2.5 \\ 3.5 \end{bmatrix} \begin{matrix} no \\ no \\ no \end{matrix}$$

$\Lambda = \emptyset$.

Step Two

For each i$\in \Omega$, determine α_j^+ by solving $((\widetilde{A}_i)^+_{\alpha_i^+})^T$x=$(\tilde{\mathrm{b}}_i)^-_{\alpha_i^+}$):

i= 1

$(1.5 - .5\alpha)1.5 + (3.1667 - .5\alpha).5 = 3.5 + .5\alpha$, Solution is : $\alpha_1^+ = .2222$

For each i$\in$?, determine α_i^- by solving $((\widetilde{A}_i)^-_{\alpha_i^+})^T$x=$(\tilde{\mathrm{b}}_i)^+_{\alpha_i^+}$)

i= 2

$(.5 + .5\alpha)1.5 + (.5 + .5\alpha).5 = 2.5 - .5\alpha$, Solution is : $\alpha_2^- = 1.0$

i= 3

$(1.5 + .5\alpha)1.5 = 3.5 - .5\alpha$, Solution is : $\alpha_3^- = 1.0$

Step Three

Determine the gradient of EA$(\tilde{f}(x))$:

$$\partial(\mathrm{EA}(\tilde{\mathrm{f}}(\mathrm{x})))/\partial \mathrm{x}_1 = 2 - \int_0^{.2222} \tfrac{1}{2}(3.5 - .5\alpha)(1.5 - .5\alpha)d\alpha$$

$$- \int_0^1 \tfrac{1}{2}(2.5 - .5\alpha)(1.5 - .5\alpha)d\alpha - \int_0^1 \tfrac{1}{2}(3.5 - .5\alpha)(2.5 - .5\alpha)d\alpha = -3.6363$$

$$\partial(\mathrm{EA}(\tilde{\mathrm{f}}(\mathrm{x})))/\partial \mathrm{x}_2 = 1 - \int_0^{.2222} \tfrac{1}{2}(3.5 - .5\alpha)(\tfrac{8}{3} + .5 - .5\alpha)d\alpha$$

$- \int_0^1 \frac{1}{2}(2.5 - .5\alpha)(1.5 - .5\alpha)d\alpha = -1.6075$

We search for an improved solution at

$$(1.5, .5) + .05(-3.6363, -1.6075) = (1.3182, .4196) \text{ where } EA(\tilde{f}(x)) =$$

$$\int_0^1 \frac{1}{2}((2.5 - .5\alpha)1.3182 + (1.5 - .5\alpha).4196)d\alpha$$

$$- \int_0^1 \frac{1}{2}(2.5 + .5\alpha)\max[0, (.5 + .5\alpha)1.3182 + (\tfrac{8}{3} - .5 + .5\alpha).4196 - (4.5 - .5\alpha)]d\alpha$$

$$- \int_0^1 \frac{1}{2}(1.5 + .5\alpha)\max[0, (.5 + .5\alpha)1.3182 + (.5 + .5\alpha).4196 - (2.5 - .5\alpha)]d\alpha$$

$$- \int_0^1 \frac{1}{2}(2.5 + .5\alpha)\max[0, (1.5 + .5\alpha)1.3182 - (3.5 - .5\alpha)]d\alpha$$

$$+ \int_0^1 \frac{1}{2}((1.5 + .5\alpha)1.3182 + (.5 + .5\alpha).4196)d\alpha$$

$$- \int_0^1 \frac{1}{2}(3.5 - .5\alpha)\max[0, (1.5 - .5\alpha)1.3182 + (8/3 + .5 - .5\alpha).4196 - (3.5 + .5\alpha)]d\alpha$$

$$- \int_0^1 \frac{1}{2}(2.5 - .5\alpha)\max[0, (1.5 - .5\alpha)1.3182 + (1.5 - .5\alpha).4196 - (1.5 + .5\alpha)]d\alpha$$

$$- \int_0^1 \frac{1}{2}(3.5 - .5\alpha)\max[0, (2.5 - .5\alpha)1.3182 - (2.5 + .5\alpha)]d\alpha = 2.0648$$

This is an improvement over $x = (1.5, .5)$.

B. APPENDIX Formulas for Implementation of Fuzzy Linear Programming

Following are formulas that can be used to implement the gradient ascent algorithm when all coefficients in the original linear programming problem are replaced by trapezoidal fuzzy numbers of the form $(a_1, a_2, a_{3,} a_4)$ (see Kaufmann&Gupta [20]).

Let $A1, A2, A3, A4$ be crisp matrices where the entries of A_i are the i'th elements of the trapezoidal numbers that make up the fuzzy numbers in $\widetilde{A}$. For example if each coefficient in $\widetilde{A}$ is the trapezoidal number (5,6,7,8) then A_2 will be a matrix with all entries equal to the number 6. We define B_i, C_i, and D_i for i=1 to 4 in the same way.

With this representation we have the following formulas for the α-cuts of $\widetilde{A}$ (the formulas for $\widetilde{B}, \widetilde{C}$, and $\widetilde{D}$ are identical):

$$\widetilde{A}_{\alpha}^{-} = A1 + (A2 - A1) * \alpha$$

and

$$\widetilde{A}_{\alpha}^{+} = A4 + (A3 - A4) * \alpha$$

Let A be an $m \times n$ matrix.

Calculating α_i^+ and α_i^-

```
For i=1 to m
  if ( A4(i,:)*x > B1(i) )&( A3(i,:)*x < B2(i) )
    α_i^+=(B1(i)-A4(i,:)*x)/( (A3(i,:)-A4(i,:))*x - (B2(i)-B1(i)) )
  elseif A3(i,:)*x>=B2(i)
    α_i^+=1
  else
    α_i^+=0
  end
  if ( A2(i,:)*x>B3(i) )&( A1(i,:)*x<B4(i) )
    α_i^-=( B4(i)-A1(i,:)*x )/( (A2(i,:)-A1(i,:) )*x + ( B4(i)-B3(i) ) )
  elseif A1(i,:)*x>=B4(i)
    α_i^-=0
  else
    α_i^-=1
  end
end
```

Calculating EA($\tilde{f}$(x))

EA($\tilde{f}(x)$) =1/4*(C1+C2+C3+C4)'*x -

$\sum_{i=1}^{m}$[1/2 * D1(i) * (A1(i,:)*x-B4(i))

+ 1/2* (D1(i) * ((A2(i,:)-A1(i,:)) *x + B4(i) - B3(i))

+ (D2(i)-D1(i)) * (A1(i,:) *x - B4(i)) *(1/2)

+ 1/2* ((D2(i)-D1(i)) * ((A2(i,:) -A1(i,:)) *x

+ B4(i) - B3(i))) * (1/3)

- (1/2 * D1(i) * (A1(i,:)*x-B4(i)) * α_i^-

+ 1/2* (D1(i) * ((A2(i,:)-A1(i,:)) *x + B4(i) - B3(i))

+ (D2(i)-D1(i)) * (A1(i,:) *x - B4(i))) *1/2 * $(\alpha_i^-)^2$

+ 1/2* ((D2(i)-D1(i)) * ((A2(i,:) -A1(i,:)) *x

+ B4(i) - B3(i))) * 1/3 * $(\alpha_i^-)^3$)

+ 1/2 * D4(i) * (A4(i,:)*x-B1(i)) * α_i^+

+ 1/2* (D4(i) * ((A3(i,:)-A4(i,:)) *x + B1(i) - B2(i))

+ (D3(i)-D4(i)) * (A4(i,:) *x - B1(i))) *1/2 * $(\alpha_i^+)^2$

+ 1/2* ((D3(i)-D4(i)) * ((A3(i,:) -A4(i,:)) *x

+ B1(i) - B2(i))) * 1/3 * $(\alpha_i^+)^3$]

Calculating $\partial(\mathbf{EA}(\tilde{\mathbf{f}}(\mathbf{x})))/\partial\mathbf{x}_j$

$\partial(\mathrm{EA}(\tilde{\mathrm{f}}(\mathrm{x})))/\partial\mathrm{x}_j$ = 1/4*(C1(j)+C2(j)+C3(j)+C4(j)) -

$\sum_{i=1}^{m}$[1/2 * D1(i) * A1(i,j)

+ 1/2* (D1(i) * (A2(i,j)-A1(i,j))

+ (D2(i)-D1(i)) * A1(i,j)) *(1/2)

+ 1/2* ((D2(i)-D1(i)) * (A2(i,j) -A1(i,j))) * (1/3)

- (1/2 * D1(i) * A1(i,j) * α_i^-

+ 1/2* (D1(i) * (A2(i,j)-A1(i,j))

+ (D2(i)-D1(i)) * A1(i,j)) * 1/2 * $(\alpha_i^-)^2$

+ 1/2* ((D2(i)-D1(i)) * (A2(i,j) -A1(i,j))) * 1/3 * $(\alpha_i^-)^3$)

+ 1/2 * D4(i) * A4(i,j) * α_i^+

+ 1/2* (D4(i) * (A3(i,j)-A4(i,j))

+ (D3(i)-D4(i)) * A4(i,j)) *1/2 * $(\alpha_i^+)^2$

+ 1/2* ((D3(i)-D4(i)) * (A3(i,j) -A4(i,j))) * 1/3 * $(\alpha_i^+)^3$]

REFERENCES

[1] G. Beliakov, Fuzzy sets and membership functions based on probabilities, *Information Sciences 91* (1996) 95-111

[2] L. Breiman, Probability (SIAM, Philadelphia, 1992).

[3] J.J. Buckley, Joint solution to fuzzy programming problems, *Fuzzy Sets and Systems 72* (1995) 215-220.

[4] J.J. Buckley, Possibilistic linear programming with triangular fuzzy numbers, *Fuzzy Sets and Systems 26* (1988) 135-138.

[5] J.M. Cadenas and J.L. Verdegay, Using Fuzzy Numbers in Linear Programming, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, Vol. 27 No. 6 (December 1997) 1016-1022

[6] M. Delgado, J.L. Verdegay, M.A. Vila, A general model for fuzzy linear programming, *Fuzzy Sets and Systems 29* (1989) 21-29.

[7] P. Diamond, Congruence classes of fuzzy sets form a Banach space, *J. Math. Anal. Applns.*, 162 (1991), 144-151

[8] P. Diamond & P. Kloeden, Metric Spaces of Fuzzy Sets, *Fuzzy Sets and Systems* 35 (1990) 241-249.

[9] D. Dubois, H. Prade, Bayesian conditioning in possibility theory, *Fuzzy Sets and Systems 92* (1997) 223-240.

[10] D. Dubois, H. Prade, *Possibility Theory an approach to computerized processing of uncertainty* (Plenum Press, New York, 1988).

[11] D. Dubois, H. Prade, Random sets and fuzzy interval analysis, *Fuzzy Sets and Systems 42* (1991) 87-101.

[12] D. Dubois, H. Prade, The three semantics of fuzzy sets, *Fuzzy Sets and Systems 90* (1997) 141-150.

[13] R. Fuller and H.J. Zimmermann, Fuzzy reasoning for solving fuzzy mathematical programming problems,*Fuzzy Sets and Systems 60* (1993), 121-133.

[14] M. Inuiguchi, M. Sakawa, Possible and necessary optimality tests in possibilistic linear programming problems, *Fuzzy Sets and Systems 67* (1994) 29-46.

[15] K.D. Jamison and W.A. Lodwick, *Fuzzy linear programming using a penalty method* UCD/CCM Report No. 131 (1998) (accepted for publication in Fuzzy Sets and Systems)

[16] K.D. Jamison and W.A. Lodwick, *Minimizing Unconstrained Fuzzy Functions UCD/CCM Report No. 80* (1996)(accepted for publication in Fuzzy Sets and Systems)

[17] K.D. Jamison, *Possibilities as cumulative subjective probabilities and a norm on the space of congruence classes of fuzzy numbers motivated by an expected utility functional*(1998)(accepted for publication in Fuzzy Sets and Systems)

[18] B. Julien, An extension to possibilistic linear programming, *Fuzzy Sets and Systems 64* (1994) 195-206.

[19] C. Y. Jung and V. A. Pulmano, Improved Fuzzy Linear Programming Model for Structure Designs, *Computers and Structures Vol. 58 No. 3* (1996) 471-477

[20] A. Kaufmann and M.M.Gupta, *Introduction to Fuzzy Arithmetic Theory and Applications* (Van Nostrand Reinhold, New York, 1991).

[21] G. J. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic Theory and Applications (Prentice Hall Inc., New Jersey, 1995)

[22] G. J. Klir, Is there more to uncertainty than some probability theorists might have us believe?, *Int. J. General Systems, Vol. 15* (1989) 347-378.

[23] B. Kruse, J. Gebhardt and F. Klawonn, *Foundations of Fuzzy Systems* (John Wiley & Sons, 1994)

[24] A.N. Kolmogorov and S.V. Fomin, *Introductory Real Analysis* (Dover Publications, Inc., 1975)

[25] B.S. Lee and Sung Jin Cho, A fixed point theorem for contractive-type fuzzy mappings, *Fuzzy Sets and Systems* 61 (1994) 309-312.

[26] D.V. Lindley, *Bayesian Statistics, A Review* (Society for Industrial and Applied Mathematics, Montpellier, Vt, 1972)

[27] W.A. Lodwick, Analysis of structure in fuzzy linear programming, *Fuzzy Sets and Systems* 38 (1990) 15-26.

[28] W.A. Lodwick and K.D. Jamison, *Chapter 19 A Computational Method for Fuzzy Optimization*, Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Netwook Approaches (Kluwer Academic Publishers, 1997)

[29] W.A. Lodwick and K.D. Jamison, *Interval Methods and Fuzzy Optimization* (International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems Vol. 5., No. 3 (1997) 239-249)

[30] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming* (Addison-Wesley Publishing Company, Reading, MA 1972)

[31] M.K. Luhandjula, Fuzzy optimization: An appraisal, *Fuzzy Sets and Systems* 30 (1989), 257-282.

[32] R.E. Moore, *Methods and Applications of Interval Analysis* (SIAM, Philadelphia, 1979).

[33] J. R. Munkres, *Topology, A First Course* (Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1975).

[34] C. V. Negoita and D.A. Ralescu, *Applications of Fuzzy Sets to Systems Analysis* (John Wiley and Sons, New York, 1975)

[35] J. Von Neumann and O. Morgenstern, Theory of Games and Economic Behavior (3rd edition Princeton University Press, Princeton, NJ 1953)

[36] H. Rommelfanger, Fuzzy linear programming and applications, *European Journal of Operational Research 92* (1996) 512-527

[37] R. Royall, *Statistical Evidence, A likelihood paradigm* (Chapman & Hall, 1997)

[38] H.L. Royden, *Real Analysis, second edition* (The Macmillan Company, New York,1968)

[39] W. Rudin, *Principles of Mathematical Analysis* (McGraw Hill Book Company, 1976)

[40] J.J. Saade, Maximization of a function over a fuzzy domain, *Fuzzy Sets and Systems 62* (1994), 55-70.

[41] M. Sakawa and H. Yano, An interactive fuzzy satisficing method for multiobjective nonlinear programming problems with fuzzy parameters, *Fuzzy Sets and Systems 30* (1989), 221-238.

[42] M. Sasaki, Fuzzy functions, *Fuzzy Sets and Systems* 55(1993), 295-301.

[43] A. L. Soyster, Convex programming with set-inclusive constraints and applications to inexact linear programming, *Operations Research* 21(5) (1973), 1154-1157.

[44] H. Tanaka, T. Okuda and K. Asai, On fuzzy mathematical programming, *J. of Cybernet.* 3 (1974), 37-46.

[45] J.L. Verdegay, Fuzzy mathematical programming, in: M.M. Gupta and E. Sanchez, Eds., *Fuzzy Information and Decision Processes* (North-Holland, Amsterdam, 1982), 231-237.

[46] Z. Wang and G.J. Klir, *Fuzzy Measure Theory* (Plenum Press, New York, 1992).

[47] B. Werners, Interaktive Entscheidungsunterstützung durch ein flexibles mathematicshches Programmierungssytem, Minerva Publikation, München, 1984.

[48] L.A. Zadeh, Fuzzy Sets, Infor. Control 8 (1965) 338-353

[49] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems 1* (1978) 3-28.

[50] D. Zhang and C. Guo, Fuzzy integrals of set-valued mappings and fuzzy mappings, *Fuzzy Sets and Systems* 75 (1995) 103-109.

[51] H.J. Zimmermann, Fuzzy mathematical programming, *Comput. and Oper. Res.* 10 (1983), 291-298.

[52] H.J. Zimmermann, *Fuzzy Sets and Applications*, Kluwer Nijhoff, Dordrecht, 1985.