Chapter 2
Monotone Measure Probability Representations and Weighted Fuzzy Statistics

Abstract The basic definitions and propositions of the well-known Campos–Bolanos probability representations (CBR) and Murofushi–Sugeno probability representations (MSR) of several monotone measures are presented. These include Choquet capacities of order two (in general), Dempster–Shafer belief structure, Sugeno λ-measures, and possibility measure.

Connections between the CBR and MSR are considered. Analogously to CBR, theorems of existence of Choquet capacities of order two are proved for MSR. A definition of a distance between monotone measures in MSR is introduced, which is similar to the definition of a distance for CBR.

Some properties of monotone expectation in relation to MSR are also obtained.

On the basis of CBR of a monotone measure, its restoring problem is considered using insufficient expert data.

Most typical weighted expected values on insufficient expert data are defined based on the CBR.

2.1 On Campos–Bolanos Representations (CBR) of Monotone Measures

2.1.1 Introduction

Sometimes, it is more appropriate to consider nonadditive but monotone valuation to express human subjectivity [2, 8, 9, 13, 17–19, 27, 39, 40, 44, 49, 63, 71, 72, 93, 98, 110, 118, 119, 138, 139, 141, 157, 163, 179, 181, and others], for example monotone (fuzzy) measure, in which nonadditivity separates it from important properties of the probability measure. It is clear that nonadditivity limits the possibilities of practical applications of monotone measures. Therefore, researchers continue to study probability representations of monotone measures, providing new perspectives for their use [2, 8, 9, 49, 93, 130, 135, 136, 138, 139, 181, and others].

G. Sirbiladze, *Extremal Fuzzy Dynamic Systems: Theory and Applications*, IFSR International Series on Systems Science and Engineering 28,
DOI 10.1007/978-1-4614-4250-9_2, © Springer Science+Business Media, LLC 2013

From this point of view, we consider two probability representations of monotone measures on a finite set: the Campos–Bolanos [8] representation (CBR) and the Murofushi–Sugeno [93] representation (MSR).

In Sect. 2.1.2, basic definitions of the well-known Campos–Bolanos and Murofushi–Sugeno probability representations for monotone measures and principal theorems are considered. The relation between CBR and MSR for monotone measures is constructed. It is presented schematically and compositionally. A class of monotone measures, Choquet second-order capacities, is considered. The properties of Choquet capacities of order two in MSR have been studied analogously to what was done for CBR. Some properties of the monotone expectation in connection with MSR have been also studied. The notion of distance on the class of monotone measures in MSR is introduced, which is analogous to the corresponding definition in CBR. By parameterization, it takes an "almost" clear form. The correctness principle between two monotone measures at the corresponding dual measure is preserved. The example of a two-element set is considered.

A priori expert information can be represented in different ways. Below, we consider the representations of this information in terms of some monotone measure based on expert information (namely, Sugeno λ-measures or monotone measures associated with a Dempster–Shafer belief structure).

Axiomatic or eristic approaches to the discrete Choquet integral and Sugeno integral were presented in [8, 28, 42, 45, 56, 57, 72, 86, 87, 93, 106, 134–136, 154, and others] as an aggregation instrument for criteria construction.

We consider different CBR for Dempster–Shafer belief structure, namely, consonant, dissonant, and combined bodies of evidence.

Derived from the body of evidence, we further consider the associated belief and plausibility measures, possibility measure, and all other monotone measures associated with a belief structure. We represent all these monotone measures by their associated probabilities and connect them to the focal probabilities of the body of evidence. Also, fuzzy expected value (FEV) and monotone expectation (ME) CBR are constructed as an important defuzzification instrument in the decision-making environment.

2.1.2 A Monotone Measure on a Finite Set and a Monotone Expectation: Campos–Bolanos Probability Representation of a Monotone Measure

There are two classical approaches to the study of inexact data. When experimental data are sufficiently exact, probabilistic–statistical methods are used to process them and estimate their general characteristics. If data presented are vague, inexact, and have intervals, then the methods of the theory of errors can be successfully applied.

However, there are cases in which neither the probabilistic–statistical methods nor those of the theory of errors provide satisfactory results. Then one has to investigate the nature of the means (description, measurement, scaling, etc.) by which this information was received.

If data are represented in intervals, their distribution is obscure, or they overlap and are described or obtained by an individual expert (insufficient expert data), then they are considered to be of a combined nature. In that case, along with probabilistic–statistical uncertainty, there arises so-called possibilistic uncertainty produced by an individual (expert), which demands the application of fuzzy analysis methods. In such situations, only probabilistic–possibilistic analysis can provide satisfactory results through the use of the fuzzy methods to be discussed below.

In trying to describe such insufficient expert data functionally, in many real situations the property of additivity remains unrevealed for a measurable representation of a set, and this creates an additional restriction. Hence, to study such data it is frequently better to use monotone estimators instead of additive ones.

We introduce the definition of a monotone measure [8] adapted to the case of a finite referential.

Definition 2.1. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set and g a set function

$$g : \mathscr{B}(X) \to [0,1], \tag{2.1}$$

where $\mathscr{B}$ is the power set of X.

1. We will say that g is a monotone measure on X if it satisfies

$$\begin{aligned} &\text{(i)}\ \ g(\varnothing) = 0; \quad g(X) = 1. \\ &\text{(ii)}\ \ \forall A, B \subseteq X, \ \text{ if } A \subseteq B, \text{ then } g(A) \leq g(B). \end{aligned} \tag{2.2}$$

2. $(X, \mathscr{B}(X), g)$ is called a monotone measure space.

A monotone measure is a normalized and monotone set function. It can be considered an extension of the probability concept, where additivity is replaced by the weaker condition of monotonicity.

Nonadditive but monotone measures were first used in fuzzy analysis in the 1980s [49, 141, and others].

The fuzzy integral is a functional that assigns some number or a compatibility value to each fuzzy subset when the monotone measure is already fixed. As known [38, 49, 141, and others], the concept of a fuzzy integral condenses the information provided by a compatibility function of a fuzzy set and a monotone measure. Having the monotone measure determined, we can estimate a fuzzy subset by the most typical compatibility value (MTV or a fuzzy average [38, Sect. 1.6]). The MTV is essentially different in content and significance from a probabilistic average even when a probabilistic measure is used instead of a monotone measure. The preimage of the MTV with respect to a compatibility function distinguishes it from the universe of most typical representative values of the considered fuzzy subset.

As already known, fuzzy averages (MTVs) differ both in form and content from probabilistic-statistical averages and other numerical characteristics such as *mode* and *median*. Nevertheless, in some cases, "nonfuzzy" (objective) and "fuzzy" (subjective) averages coincide [31, 38], and so on. For a given set of fuzzy subsets with compatibility function values from the interval $[0,1]$, the fuzzy average determines the most typical representative compatibility value ME—monotone expectation.

The following fuzzy integral (based on the Choquet operator [13]) is the monotone expectation, which was defined by Bolaňos et. al. [8]:

Definition 2.2. Let g be a monotone measure on X and $h : X \to \mathbb{R}_0^+$ a nonnegative function. The monotone expectation of h with respect to g is

$$E_g(h) = \int_0^{+\infty} g(H_\alpha)\,d\alpha, \tag{2.3}$$

where $H_\alpha = \{x \in X \mid h(x) \geq \alpha\}$.

The monotone expectation always exists and is finite for each g and h. It is obvious that $E_g(\cdot)$ is a generalization of the mathematical expectation: that is what it becomes when the monotone measure used is a probability measure,

$$E_P(h) = \int_X h\,dP, \tag{2.4}$$

where P denotes a probability measure, E_P the mathematical expectation.

Since the monotone expectation is a generalization of the mathematical expectation, it can be questioned whether the former possesses some weaker property in relation to additivity than the latter. The following proposition gives an expression of the monotone expectation that permits us to analyze this question.

Proposition 2.1 ([8]). *If the values of a nonnegative function h are ordered as*

$$h(x_1) \leq h(x_2) \leq \cdots \leq h(x_n),$$

then the monotone expectation of h with respect to a monotone measure g can be written as

$$E_g(h) = \sum_{i=1}^{n} h(x_i)(g(A_i) - g(A_{i+1})), \tag{2.5}$$

where $A_i = \{x_i, x_{i+1}, \ldots, x_n\}$, $i = 1, \ldots, n$, $g(A_{n+1}) = 0$.

Thus, the monotone expectation is an additive functional for functions ordered monotonically.

We can also notice that $E_g(h)$ is an average of the h function values weighted by

$$p_i = g(A_i) - g(A_{i+1}),\;\; i = 1, \ldots, n,\;\; p_n = g(A_n).$$

Since

$$\sum_{i=1}^{n} p_i = g(A_1) = g(X) = 1 \quad \text{and} \quad p_i \geq 0, \quad i = 1, \ldots, n,$$

the values p_i can be interpreted as the values of a probability function. Then $E_g(h)$ is equivalent to the mathematical expectation of h with respect to that probability distribution.

The values p_i depend on the monotone measure g and the sets A_i, which in general depend on ordering of the values of h. So we may state the following result.

Proposition 2.2. *The monotone expectation of a nonnegative function h with respect to a monotone measure g coincides with the mathematical expectation of h with respect to a probability measure that depends only on g and the ordering of the values of h.*

Following Proposition 2.1, the maximum number of probability distributions coincides with the number of possible orderings or permutations in a set with n elements, that is, $n!$.

Thus, it makes sense to associate the $n!$ probabilities to each monotone measure, provided that they are deduced from this monotone measure through the different possible orderings.

In general, the possible orderings of the elements of X are given by the permutations of a set with n elements, which form the group S_n.

Definition 2.3 ([8]). The probability functions $\{P_\sigma\}_{\sigma \in S_n}$ defined by

$$\begin{aligned} P_\sigma(x_{\sigma(1)}) &= g(\{x_{\sigma(1)}\}), \ldots, \\ P_\sigma(x_{\sigma(i)}) &= g(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) - g(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}), \ldots, \\ P_\sigma(x_{\sigma(n)}) &= 1 - g(\{x_{\sigma(1)}, \ldots, x_{\sigma(n-1)}\}), \end{aligned} \tag{2.6}$$

for each $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n)) \in S_n$, are called the associated probabilities class (APC) of the monotone measure g.

An interesting case is that in which the monotone measure is a probability; in such a case, all associated probabilities are equal:

Proposition 2.3 ([8]). 1. *A monotone measure g is a probability measure if and only if its $n!$ associated probabilities coincide.*

2. $\forall A \in X$*:* $\exists \sigma = \sigma_A \in S_n$ *such that* $g(A) = P_{\sigma_A}(A)$.

Definition 2.4 ([8]). $g_*, g^* : \mathscr{B}(X) \to [0,1]$ are called dual monotone measures if $\forall A \in \mathscr{B}(X) \;\; g_*(A) = 1 - g^*(\overline{A})$.

The concept of duality is very important, since it permits one to obtain alternative representations of a piece of information. So, we will consider a monotone measure and its dual measure to contain the same information, but codified in different ways.

The most remarkable case in which different monotone measures provide the same $n!$ probabilities, but ordered in different ways, is the case of dual monotone measures. Before presenting it in the following proposition, we need a definition:

Definition 2.5 ([8]). We will say that two permutations $\sigma, \sigma^* \in S_n$ are dual if

$$\sigma^*(i) = \sigma(n-i+1), \quad i = 1,\ldots,n.$$

Proposition 2.4 ([8]). *A necessary and sufficient condition for two monotone measures g_* and g^* to be dual is that they have the same $n!$ associated probabilities corresponding to dual permutations, that is, $P_{*\sigma}(x_{\sigma(i)}) = P^*_{\sigma^*}(x^*_{\sigma(i)})$, if σ and σ^* are dual.*

If we accept that a monotone measure and its dual measure contain the same information, but codified in different ways, then the above-mentioned result could be interpreted by saying that the $n!$ associated probabilities contain the information and that the different orderings are different codifications of this information.

Let $S_n^{(h)}$ $(S_n^{(h)} \subset S_n)$ be the subgroup of all permutations such that $\forall \sigma \in S_n^{(h)}$

$$h(x_{\sigma(1)}) \leq h(x_{\sigma(2)}) \leq \cdots \leq h(x_{\sigma(n)}). \tag{2.7}$$

Following Proposition 2.1 and Definition 2.3, the monotone expectation with respect to a monotone measure can be written as mathematical expectation with respect to any associated probability P_σ $(\sigma \in S_n^{(h)})$,

$$\begin{aligned}
E_{g_*}(h) &= E_{P_{*\sigma}}(h) = \sum_{i=1}^{n} P_{*\sigma}(x_{\sigma(i)})h(x_{\sigma(i)}) = \sum_{j=1}^{n} P_{*\sigma}(x_j)h(x_j),\\
E_{g^*}(h) &= E_{P^*_\sigma}(h) = \sum_{i=1}^{n} P^*_\sigma(x_{\sigma(i)})h(x_{\sigma(i)})\\
&= \sum_{j=1}^{n} P^*_\sigma(x_j)h(x_j) = \sum_{j=1}^{k} P_{*\sigma^*}(x_j)h(x_j),
\end{aligned} \tag{2.8}$$

where $\{P_{*\sigma}\}_{\sigma\in S_n} = \{P^*_\sigma\}_{\sigma\in S_n}$ are the classes of associated probabilities for g_* and g^* monotone measures, respectively.

Now we consider some concrete classes of monotone measures and their probability representations.

2.1.3 *Choquet Capacities of Order Two in CBR*

An especially interesting class of monotone measures are Choquet capacities of order two [13], because they cover a great number of monotone measures.

Definition 2.6 ([8]). Let (g_*, g^*) be a pair of dual monotone measures. Then g_* is a lower capacity of order two if

$$\forall A, B \subseteq X, \quad g_*(A \cup B) + g_*(A \cap B) \geq g_*(A) + g_*(B),$$

and g^* is an upper capacity of order two if

$$\forall A, B \subseteq X, \quad g^*(A \cup B) + g^*(A \cap B) \leq g^*(A) + g^*(B).$$

The most frequently used classes of monotone measures such as belief and plausibility measures [18, 67, 68, 70, 73, 83, 83, 169, 170, and others], necessity and possibility measures [23, 28, 83, and others], λ-measures [141], and probabilities are capacities of order two.

Proposition 2.5 ([8]). *Let (g_*, g^*) be a pair of dual monotone measures. Then g_* is a lower capacity of order two $(g^*$ is an upper capacity of order two$)$ if and only if*

$$\begin{aligned} g_*(A) &= \min_{\sigma \in S_n} P_\sigma(A) \quad \forall A \subseteq X, \\ (g^*(A) &= \max_{\sigma \in S_n} P_\sigma(A) \quad \forall A \subseteq X).L \end{aligned} \tag{2.9}$$

So the main characteristic of a capacity of order two is that it depends only on the probabilities associated to such a measure, but does not depend on the permutations that generate them: we can regenerate the initial monotone measure knowing only its associated probabilities, without having to know the corresponding permutations. This characteristic makes the use of capacities of order two by means of associated probabilities especially easy.

From this property, the following result is evident and valid for every monotone measure:

Proposition 2.6 ([8]). *If P_σ, $\sigma \in S_n$, are the probabilities associated with a monotone measure g, then for every $h : X \to \mathbb{R}_0^+$, we have*

$$\min_{\sigma \in S_n} E_{P_\sigma}(h) \leq E_g(h) \leq \max_{\sigma \in S_n} E_{P_\sigma}(h). \tag{2.10}$$

Proposition 2.7 ([8]). *A necessary and sufficient condition for a pair of dual monotone measures (g_*, g^*) to be lower and upper capacities of order two respectively is that $\forall h : X \to \mathbb{R}_0^+$,*

$$E_{g_*}(h) = \min_{\sigma \in S_n} E_{P_\sigma}(h), \quad E_{g^*}(h) = \max_{\sigma \in S_n} E_{P_\sigma}(h). \tag{2.11}$$

2.1.4 Fuzzy Expected Value (FEV): The Campos–Bolanos Probability Representation

In this subsection, we discuss the main estimator of fuzzy statistics: the FEV of a population. The FEV determines the MTV for a compatibility function [56, 57, 134, 135].

Let h be a compatibility function of some fuzzy subset of X.

Definition 2.7. The FEV of the compatibility function h with respect to the monotone measure g is Sugeno's integral over X:

$$FEV_g(h) = \int_X h \circ g(\cdot) \equiv \sup_{\alpha \in [0,1]} \{\alpha \wedge g(H_\alpha)\}, \tag{2.12}$$

where $\wedge$ denotes the minimum of two arguments.

It clearly follows that the FEV somehow "averages" the values of the compatibility function h not in the sense of a statistical average but by cutting off subsets of α level whose values of the monotone measure g are either sufficiently "high" or sufficiently "low."

Thus the FEV gives a concrete value of the compatibility function h, this value being the most typical characteristic of all possible values with respect to the monotone measure g, obtained by cutting off the "upper" and "lower" strips on the graph of $g(H_\alpha)$.

Thus the information carried by h and g is condensed in the FEV, which is the most typical value of all compatibility levels of h.

Proposition 2.8 ([38]). *If the values of a compatibility function h are ordered as*

$$h(x_1) \le h(x_2) \le \cdots \le h(x_n),$$

then the FEV of h with respect to a monotone measure g can be written as

$$FEV_g(h) = \bigvee_{i=1}^{n} \{h(x_i) \wedge g(A_i)\} = \bigwedge_{i=1}^{n} \{h(x_i) \vee g(A_i)\}, \tag{2.13}$$

where $A_i = \{x_i, x_{i+1}, \ldots, x_n\}$, $i = 1, \ldots, n$, *and where* $\vee$ *is the maximum of two arguments.*

Let $S_n^{(h)}$ ($S_n^{(h)} \subset S_n$) be the subgroup of all permutations such that $\forall \sigma \in S_n^{(h)}$

$$h(x_{\sigma(1)}) \le h(x_{\sigma(2)}) \le \cdots \le h(x_{\sigma(n)}).$$

Following Proposition 2.8, the FEV can be written by the associated probabilities of a monotone measure g as

$$FEV = \bigvee_{i=1}^{n} \{h(x_{\sigma(i)}) \wedge g(A_i^{(\sigma)})\} = \bigwedge_{i=1}^{n} \{h(x_{\sigma(i)}) \vee g(A_i^{(\sigma)})\}, \tag{2.14}$$

where $A_i^{(\sigma)} = \{x_{\sigma(i)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}\}$, $i = 1, 2, \ldots, n$.

Let (g_*, g^*) be a pair of dual lower and upper capacities of order two. Using Proposition 2.7 and formula (2.14), the FEV can be written as follows: $\forall \sigma \in S_n^{(h)}$,

$$\begin{aligned} FEV_{g_*}(h) &= \bigwedge_{i=1}^{n} \bigwedge_{\sigma' \in S_n} \{h(x_{\sigma(i)}) \vee P_{*\sigma'}(A_i^{(\sigma)})\}, \\ FEV_{g^*}(h) &= \bigvee_{i=1}^{n} \bigvee_{\sigma' \in S_n} \{h(x_{\sigma(i)}) \wedge P^*_{\sigma'}(A_i^{(\sigma)})\} \\ &= \bigvee_{i=1}^{n} \bigvee_{\sigma' \in S_n} \{h(x_{\sigma(i)}) \wedge P_{*\sigma'^*}(A_i^{(\sigma)})\}, \end{aligned} \tag{2.15}$$

where $\{P_{*\sigma'}\}_{\sigma' \in S_n} = \{P^*_{\sigma'}\}_{\sigma' \in S_n}$ are equal classes of associated probabilities for g_* and g^* monotone measures, respectively.

2.1.5 Dempster–Shafer's Belief Structure and Its CBR

The theory of evidence (Dempster–Shafer belief structure) [18, 67, 68, 70, 73, 83, 169, 170, and others] is a powerful tool that enables one to build (1) models of decisions and their risks in uncertain environments; (2) optimization criteria in general uncertain environments.

The theory of evidence is based on two dual monotone measures: belief measures and plausibility measures. These classes of monotone measures are subclasses of classes of dual lower and upper capacities of order two. This is easy to prove after the introduction of belief and plausibility measures, which can be characterized by the set function

$$m : \mathscr{B}(X) \to [0,1],$$

which is required to satisfy two conditions:

$$\begin{aligned} &\text{(a)} \quad m(\varnothing) = 0, \\ &\text{(b)} \quad \textstyle\sum_{B \in \mathscr{B}(X)} m(B) = 1. \end{aligned} \tag{2.16}$$

This function is called a basic probability assignment (B.P.A.). For each set $B \in \mathscr{B}(X)$, the value $m(B)$ expresses the proportion that all available and relevant evidence supporting the claim that a particular element of X, whose characterization in terms of relevant attributes is deficient, belongs to the set B. This value, $m(B)$, pertains solely to one set, namely the set B; it does not imply any additional claims regarding subsets of B. If there is some additional evidence supporting the claim that the element belongs to a subset of B, say $B_1 \subseteq B$, it must be expressed by another value $m(B_1)$ [73].

Let m be a P.B.A. on X. The plausibility measure Pl associated with m is given by

$$Pl(A) = \sum_{B: A \cap B \neq \varnothing} m(B), \quad \forall A \in \mathscr{B}(X), \tag{2.17}$$

and the belief measure Bel associated with m is given by

$$Bel(A) = \sum_{B:B\subset A} m(B), \quad \forall A \in \mathscr{B}(X). \tag{2.18}$$

Inverse procedures are also possible. Given, for example, a belief measure Bel, the corresponding basic probability assignment m is determined for all $A \in \mathscr{B}(X)$ by the formula

$$m(A) = \sum_{B:B\subseteq A} (-1)^{|A\setminus B|} Bel(B), \tag{2.19}$$

where $|A \setminus B|$ is the cardinality of the set difference of A and B.

Given a B.P.A., every set $A \in \mathscr{B}(X)$ for which $m(B) > 0$ is called a focal element. The pair $\langle \mathscr{F}, m\rangle$, where $\mathscr{F}$ denotes the set of all focal elements induced by m, is called a body of evidence.

Because Bel is a lower capacity of order two, then using Proposition 2.7 and formulas (2.18), (2.19), we obtain a probability representation of the B.P.A. $\forall A \in \mathscr{B}(X)$:

$$\begin{aligned} \min_{\sigma\in S_n} P_\sigma^{(Bel)}(A) &= \sum_{B\in\mathscr{F}:B\subseteq A} m(B), \\ m(A) &= \sum_{B\in\mathscr{F}:B\subseteq A} (-1)^{|A\setminus B|} \min_{\sigma\in S_n} P_\sigma^{(Bel)}(B), \end{aligned} \tag{2.20}$$

where $\{P_\sigma^{(Bel)}\}_{\sigma\in S_n}$ is the associated probabilities class of the monotone measure Bel.

2.1.6 Possibility Measure and Its CBR

When the focal elements of a body of evidence $\langle \mathscr{F}, m\rangle$ are required to be nested, $\mathscr{F} = \{A_{j_1} \subset A_{j_2} \subset \cdots \subset A_{j_k}\}$, the associated belief and plausibility measures are called consonant [73]. The special branch of evidence theory that deals only with bodies of evidence whose focal elements are nested (consonant body of evidence) is referred to as possibility theory [27].

Special counterparts of Bel measures and Pl measures in possibility theory are called necessity (Nes) measures and possibility (Pos) measures, respectively.

Proposition 2.9 ([73]). *Given a consonant body of evidence $\langle \mathscr{F}, m\rangle$, the associated consonant belief (necessity) and plausibility (possibility) measures possess the following properties:*

$$\begin{aligned} \mathrm{Nes}(A\cap B) &= \mathrm{Nes}(A) \wedge \mathrm{Nes}(B) \ \ \textit{for all}\ \ A, B \in \mathscr{B}(X), \\ Pos(A\cup B) &= Pos(A) \vee Pos(B) \ \ \textit{for all}\ \ A, B \in \mathscr{B}(X). \end{aligned} \tag{2.21}$$

Proposition 2.10 ([73]). *Every possibility measure Pos on $\mathscr{B}(X)$ can be uniquely determined by its possibility distribution function $\rho : X \to [0,1]$ via the formula $\forall A \in \mathscr{B}(A)$*

$$Pos(A) = \max_{x \in A} \rho(x). \tag{2.22}$$

Assume that the finite universe $X = \{x_1, x_2, \ldots, x_n\}$ is given and let $A_{j_1} \subset A_{j_2} \subset \cdots \subset A_{j_k}$, where $A_i = \{x_1, x_2, \ldots, x_{j_i}\}$ for $i = 1, 2, \ldots, k$, be some consonant body of evidence: $\mathscr{F} = \{A_{j_1} \subset A_{j_2} \subset \cdots \subset A_{j_k}\}$. Let $m_{j_l} \equiv m(A_{j_l})$, $l = 1, \ldots, k$, $\rho_i \equiv \rho(x_i)$, $i = 1, \ldots, n$, $\rho_1 \equiv 1$. Then we have the k-tuple

$$m = \langle m_{j_1}, m_{j_2}, \ldots, m_{j_k} \rangle \tag{2.23}$$

and n-tuple

$$\rho = \langle \rho_1, \rho_2, \ldots, \rho_n \rangle. \tag{2.24}$$

It is easy to show that

$$\begin{cases} \rho_i = \sum\limits_{l: x_i \in A_{j_l} \in \mathscr{F}} m_{j_l}, \quad i = 1, 2, \ldots, n, \\ \\ m_{j_l} = \rho_{j_l} - \rho_{j_{l+1}}, \quad \rho_{j_{k+1}} \equiv 0, \quad l = 1, 2, \ldots, k. \end{cases} \tag{2.25}$$

Let $\{P_\sigma^{(Pos)}\}_{\sigma \in S_n}$ be the associated probabilities class of a possibility measure Pos. Then we have the following connection among $\{\rho_i\}$, $\{m_{j_l}\}$, and $\{P_\sigma\}_{\sigma \in S_n}$: $\forall \sigma \in S_n$

$$\begin{aligned} P_\sigma^{(Pos)}(x_{\sigma(i)}) &= Pos(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) - Pos(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}) \\ &= \bigvee_{l=\overline{1,i}} \rho(x_{\sigma(l)}) - \bigvee_{l=\overline{1,i-1}} \rho(x_{\sigma(l)}) = \bigvee_{l=\overline{1,i}} \rho_{\sigma(l)} - \bigvee_{l=\overline{1,i-1}} \rho_{\sigma(l)} \\ &= \bigvee_{l=\overline{1,i}} \sum_{q: x_{\sigma(l)} \in A_{j_q} \in \mathscr{F}} m_{j_q} - \bigvee_{l=\overline{1,i-1}} \sum_{q: x_{\sigma(l)} \in A_{j_q} \in \mathscr{F}} m_{j_q} \\ &= \begin{cases} 0, \\ \sum\limits_{q: x_{\sigma(i)} \in A_{j_q} \in \mathscr{F}} m_{j_q} - \sum\limits_{\substack{q: x_{\sigma(i')} \in A_{j_q} \in \mathscr{F} \\ \sigma(i') < \sigma(i)}} m_{j_q}. \end{cases} \end{aligned} \tag{2.26}$$

On the other hand,

$$\rho_i = Pos(\{x_i\}) = \bigvee_{\sigma \in S_n} P_\sigma^{(Pos)}(x_i), \quad i = 1, 2, \ldots, n. \tag{2.27}$$

Since Pos is a capacity of order two, we obtain

$$m_{j_l} = \rho_{j_l} - \rho_{j_{l+1}} = \bigvee_{\sigma \in S_n} P_\sigma^{(Pos)}(x_{j_l}) - \bigvee_{\sigma \in S_n} P_\sigma^{(Pos)}(x_{j_{l+1}}), \quad l = 1, \ldots, k. \tag{2.28}$$

2.1.7 Monotone Measure Associated with a Belief Structure and Its CBR

Let m be a B.P.A. with set of all focal elements $\mathscr{F} = \{A_1, A_2, \ldots, A_q\}$. For each focal element A_j, $j = 1, \ldots, q$, let W_j be a weighting vector of dimension $|A_j|$ whose components $w_j(i)$ ($W_j \equiv \langle w_j(1), \ldots, w_j(|A_j|)\rangle$) satisfy the conditions $w_j(i) \in [0,1]$ and $\sum_{i=1}^{|A_j|} w_j(i) = 1$. We shall call these the allocation vectors. In [167] it was shown that a set function $g : \mathscr{P}(X) \to [0,1]$ defined by

$$g(A) = \sum_{j=1}^{q} \left[m(A_j) \cdot \sum_{i=1}^{|A_j \cap A|} w_j(i) \right] \tag{2.29}$$

is a monotone measure (associated with the belief structure). Thus by selecting a collection $W = \langle W_1, W_2, \ldots, W_q \rangle$ of allocation vectors, we can define a unique monotone measure associated with a belief structure.

If all the W_j are such that $w_j(1) = 1$, then the resulting monotone measure is the plausibility measure Pl. If all W_j are selected such that $w_j(|A_j|) = 1$, then this results in the belief measure Bel.

We have the following important proposition concerning all associated monotone measures with a belief structure.

Proposition 2.11 ([166]). *If g is any monotone measure generated from a collection of allocation vectors defined by (2.29), then for all $A \in \mathscr{B}(X)$, the following hold:*

(a) $Bel(A) \le g(A) \le Pl(A)$.
(b) *The Shapley's entropy of generated monotone measures coincide:*

$$E_{\text{Shapley}}(Bel) = E_{\text{Shapley}}(g) = E_{\text{Shapley}}(Pl). \tag{2.30}$$

That is, associated monotone measures with a belief structure have the same information but codified in different ways.

Now we shall compute the associated probabilities of a monotone measure g associated with the belief structure: $\forall \sigma \in S_n$, $\forall i = 1, 2, \ldots, n$

$$\begin{aligned} P_\sigma^{(g)}(x_{\sigma(i)}) &= g(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) - g(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}) \\ &= \sum_{j=1}^{q} m(A_j) \left[\sum_{l=1}^{|A_j \cap \{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}|} W_j(l) - \sum_{l=1}^{|A_j \cap \{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}|} W_j(l) \right] \\ &= \sum_{A_j \in \mathscr{F} : A_j \cap \{x_{\sigma(i)}\} \neq 0} m(A_j) W_j(|A_j \cap \{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}|). \end{aligned} \tag{2.31}$$

2.1.8 Sugeno's λ-Additive Monotone Measures and Their Probability Representation

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite reference set, $\mathscr{B}(X)$ the power set of X.

Definition 2.8 ([141]). A monotone measure $g_\lambda : \mathscr{B}(X) \to [0,1]$ $(\lambda > -1)$ is a λ-additive monotone measure if for every $A, B \in \mathscr{B}(X)$, $A \cap B = \varnothing$,

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) \cdot g_\lambda(B). \tag{2.32}$$

It is easy to verify that for every $A \in \mathscr{B}(X)$,

$$g_\lambda(A) = \frac{1}{\lambda}\left\{\prod_{x_i \in A}(1 + \lambda g_i) - 1\right\}, \tag{2.33}$$

where $0 < g_i \equiv g(\{x_i\})$, $i = 1, \ldots, n$; $\lambda > -1$ is a parameter with the following normalization condition:

$$\frac{1}{\lambda}\left\{\prod_{x_i \in X}(1 + \lambda g_i) - 1\right\} = 1. \tag{2.34}$$

Note that g_0 $(\lambda = 0)$ is a probability measure if $\sum\limits_{x_i \in X} g_i = 1$.

It is easy to prove that the λ-additive monotone measure g_λ is a capacity of order two. It is easy to verify that $g^*_\lambda = g_{-\lambda/(1+\lambda)}$. Let $\{g_i\}$, $i = 1, \ldots, n$, denote the "fuzzy weights" of singletons for the λ-additive monotone measure g.

Due to (2.33), (2.34), and (2.6) we can write the class of associated probabilities for the λ-additive monotone measure g_λ for any $\sigma \in S_n$ as

$$P_\sigma(x_{\sigma(i)}) = g_\lambda(\{x_{\sigma(i)}\}) \prod_{j=1}^{i-1}(1 + \lambda g_\lambda(\{x_{\sigma(j)}\})), \tag{2.35}$$

or more exactly, as

$$P_\sigma(x_i) = g_\lambda(\{x_i\}) \prod_{j=1}^{i(\sigma)-1}(1 + \lambda g_\lambda(\{x_{\sigma(j)}\})), \tag{2.36}$$

where $i = 1, 2, \ldots, n$, $\sigma \in S_n$; $i(\sigma)$ is the location of x_i in the permutation σ (if $i(\sigma) = 1$, then $\prod\limits_{j=1}^{0} \equiv 1$).

2.1.9 Conclusion

In this section we have represented different types of monotone measures through their associated probabilities. These representations will be further employed in the aggregation instruments of the problems of identification (restoration) of monotone measures and the construction of weighted fuzzy averages.

2.2 On the Murofishi–Sugeno Representation (MSR) of Monotone Measures

Now we consider the second important probability representation of a monotone measure, MSR [93, 117, 138, 139, and others].

Let $(\Theta, \mathscr{B}(\Theta), \lambda)$ be a finite probability measure space, where Θ is a finite set of some definite "indices" [93], and let $\eta : \mathscr{B}(X) \to \mathscr{B}(\Theta)$ be a 0–1 order-preserving homomorphism such that $g = \lambda \circ \eta$, i.e., $\eta(\varnothing) = \varnothing$, $\eta(X) = \Theta$; if $C, B \in \mathscr{B}(X)$, $C \subset B$, then $\eta(C) \subset \eta(B)$; and $\forall A \in \mathscr{B}(X)$

$$g(A) = \lambda(\eta(A)).$$

Definition 2.9 ([93]). A representation $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$ is called the Murofushi–Sugeno representation (MSR) of a monotone measure g if $\forall A \in \mathscr{B}(X)$

$$g(A) = \lambda(\eta(A)) = \sum_{\theta \in \eta(A)} \lambda_\theta \equiv \sum_{\theta \in \eta(A)} \lambda(\theta). \tag{2.37}$$

It is clear that the space $(\Theta, \mathscr{B}(\Theta))$ is not unique, but for arbitrary MSR, $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$, there exists its equivalent representation $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X)$, where Θ_X and η_X do not depend on λ and η.

Definition 2.10 ([93]). The set Θ_X is called the set of all semifilters in $\mathscr{B}(X)$, where a semifilter in $\mathscr{B}(X)$ is a collection of subsets S of $\mathscr{B}(X)$ with the following properties: $X \in S$, $\varnothing \notin S$, and if $A \in S$ and $A \subset B$, then $B \in S$.

Let η_X be a mapping from $\mathscr{B}(X)$ to $\mathscr{B}(\Theta_X)$ given by

$$\eta_X(A) = \{S \in \Theta_X \mid A \in S\}, \quad \forall A \in \mathscr{B}(X). \tag{2.38}$$

Obviously, η_X is a 0–1 order-preserving homomorphism and $\forall A \in \mathscr{B}(X)$

$$g(A) = \lambda_X(\eta_X(A)). \tag{2.39}$$

We shall now construct a λ_X-probability measure.

Definition 2.11 ([93]). MSRs $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$ and $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X)$ are said to be equivalent if $\exists T : \mathscr{B}(X) \to \mathscr{B}(\Theta_X)$ such that

$$\begin{aligned} T(\eta^+(A)) &= \eta_X^+(A), \quad \forall A \in \mathscr{B}(X), \\ \lambda^+(E) &= \lambda_X^+(T(E)), \quad \forall E \in \mathscr{B}(\Theta), \end{aligned} \tag{2.40}$$

where $\forall A \in \mathscr{B}(X)$,

$$\eta^+(A) = \eta(A) \cap \{\theta \in \Theta \mid \lambda(\theta) > 0\} \equiv \eta(A) \cap \Theta^+,$$

$$\eta_X^+(A) = \eta_X(A) \cap \{\theta \in \Theta \mid \lambda_X(\theta) > 0\} \equiv \eta_X(A) \cap \Theta_X^+.$$

We have analogous definitions for λ^+ and λ_X^+.

We note that $\forall x_i \in X,\ g(\{x_i\}) > 0$. This assumption is natural for practical purposes. It is clear that $\eta^+ = \eta$, $\eta_X^+ = \eta$, $\lambda^+ = \lambda$, $\lambda_X^+ = \lambda_X$, and $\lambda^+(B) = \lambda(B)$, $\lambda_X^+(C) = \lambda_X(C)$.

Proposition 2.12 ([93]). *For every MSR $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$ there exists its equivalent representation $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X)$ and*

$$\lambda_X(E) = \lambda(\tau^{-1}(E)), \quad \forall E \in \mathscr{B}(\Theta_X), \tag{2.41}$$

where $\tau : \Theta \to \mathscr{B}(\Theta_X)$ such that $\forall \theta \in \Theta$,

$$\tau(\theta) = \{A \in \mathscr{B}(X) \mid \theta \in \eta(A)\}. \tag{2.42}$$

It is clear that $\mathscr{B}(\Theta_X)$ are semifilters. Notice that in $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X)$, Θ_X, $\mathscr{B}(\Theta_X)$, and η_X do not depend on the monotone measure g.

Proposition 2.13 ([93]). *For every monotone measure $g : \mathscr{B}(X) \to [0,1]$ there exist a probability measure $\lambda_X : \mathscr{B}(X) \to [0,1]$ and MSR $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X)$ of g such that $\forall A \in \mathscr{B}(X)$,*

$$g(A) = \lambda_X(\eta_X(A)). \tag{2.43}$$

Considering the equivalence from Propositions 2.12 and 2.13 for the representation of g, it is sufficient to construct a λ_X-probability measure. Note that if $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda'_X)$ and $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda''_X)$ are two MSRs of g, then $\forall A \in \mathscr{B}(X)$,

$$g(A) = \lambda'_X(\eta_X(A)) = \lambda''_X(\eta_X(A)), \tag{2.44}$$

i.e., projections of the probability measures λ'_X and λ''_X on the set $\eta_X(A) \in \mathscr{B}(\Theta_X)$ coincide. Then from $[0,1]^{\mathscr{B}(\Theta_X)}$ we may select the class of probability measures of inequivalent representations of the monotone measure g:

$$L_X^g = \{\lambda_X \in [0,1]^{\mathscr{B}(\Theta_X)} \mid \forall A \in \mathscr{B}(X),\ g(A) = \lambda_X(\eta_X(A))\}. \tag{2.45}$$

Definition 2.12. The representations $\{\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X\}_{\lambda_X \in L_X^g}$ are called inequivalent representation classes (NRC) of the monotone measure g.

Notice that NRC completely describes the monotone measure g analogously to $\{P_\sigma\}_{\sigma \in S_n}$-APC in CBR.

2.2.1 Choquet Capacities of Order Two in MSR

Analogously to Sect. 2.1.2, we consider the MSR of Choquet capacities of order two.

Let

$$\{\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X\}_{\lambda_X} \in L_X^g, \quad \{\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X^*\}_{\lambda_X} \in L_X^{g^*}, \tag{2.46}$$

be NRCs respectively to the monotone measures g and g^*.

Proposition 2.14. *Measures g and g^* are dual monotone measures on $\mathscr{B}(X)$ if and only if for the pair of probability measures, we have $\forall \lambda_X \in L_X^g$, $\forall \lambda_X^* \in L_X^{g^*}$, and $\forall A \in \mathscr{B}(X)$,*

$$\lambda_X(\eta_X(A)) = \lambda_X^*(\overline{\eta_X(\overline{A})}). \tag{2.47}$$

Proof. Necessity. Let g and g^* be dual monotone measures, i.e., $\forall A \in \mathscr{B}(X)$: $g(A) = 1 - g^*(\overline{A})$. Let $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$ and $(\Theta^*, \mathscr{B}(\Theta^*), \eta^*, \lambda^*)$ be any MSRs of g and g^*, respectively. Then $\lambda(\eta(A)) = \lambda^*(\overline{\eta(\overline{A})})$. Following Proposition 2.12, $\exists \lambda_X^0 \in L_X^g$, $\lambda_X^{*0} \in L_X^{g^*}$ probability measures such that $\forall A \in \mathscr{B}(X)$,

$$\begin{aligned}\lambda_X^0(\eta_X(A)) &= \lambda_X^0(\tau^{-1}(\eta(A))) = \lambda_X^0(\eta(A)) = \lambda^{*0}(\overline{\eta^*(\overline{A})}) \\ &= \lambda_X^{*0}(\tau^{*-1}(\overline{\eta^*(\overline{A})})) = \lambda_X^{*0}(\overline{\eta_X(\overline{A})}).\end{aligned}$$

Since λ^0, λ^{*0} are any probability measures of MSR, it follows that for all measures $\lambda_X \in L_X^g$, $\lambda_X^* \in L_X^{g^*}$, there exist probability measures $\lambda \sim \lambda_X$ and $\lambda^* \sim \lambda_X^*$ such that (2.47) is true.

Sufficiency. Let (2.47) be true for all probability measures $\lambda_X \in L_X^g$, $\lambda_X^* \in L_X^{g^*}$. Consider arbitrary MSRs $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$ and $(\Theta^*, \mathscr{B}(\Theta^*), \eta^*, \lambda^*)$ corresponding to g and g^*, respectively, and their equivalent representations $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X)$ and $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X^*)$. Then $\forall A \in \mathscr{B}(X)$,

$$\begin{aligned}g(A) &= \lambda(\eta(A)) = \lambda_X(\tau^{-1}(\eta(A))) = \lambda_X(\eta_X(A)) = \lambda_X^*(\overline{\eta_X(\overline{A})}) \\ &= \lambda^*(\tau^{*-1}(\overline{\eta_X(\overline{A})})) = \lambda^*(\overline{\eta^*(\overline{A})}) = 1 - \lambda^*(\overline{\eta^*(\overline{A})}) = 1 - g^*(A).\end{aligned}$$

The proposition is proved. □

Notice that if g is an auto-dual monotone measure ($(g)^* = g$), i.e., $\forall A \in \mathscr{B}(X)$, $g(A) = 1 - g(\overline{A})$, but g is not a probability measure, then (2.47) will be changed by

$$\lambda_X(\eta_X(A)) = \lambda_X(\overline{\eta_X(\overline{A})}). \tag{2.48}$$

If g is a probability measure, then $\forall A, B \in \mathscr{B}(X)$, $A \cap B = \varnothing$,

$$\lambda_X(\eta_X(A) \cup \eta_X(B)) = \lambda_X(\eta_X(A \cup B)). \tag{2.49}$$

Analogously to Proposition 2.7 we have the following result.

Proposition 2.15. *Dual monotone measures g, g^* are Choquet lower and upper capacities of order two respectively if and only if for all probability measures $\lambda_X \in L_X^g$, $\lambda_X^* \in L_X^{g^*}$ and $\forall A \in \mathscr{B}(A)$,*

$$\lambda_X(\eta_X(A \cup B)) \geq \lambda_X(\eta_X(A) \cup \eta_X(B)),$$

$$\lambda_X^*(\eta_X(A \cup B)) \leq \lambda_X^*(\eta_X(A) \cup \eta_X(B)). \tag{2.50}$$

Proof. Necessity. Let dual monotone measures g, g^* be Choquet lower and upper capacities of order two respectively and let the MSRs of g and g^* be respectively $(\Theta_X, \mathscr{B}(\Theta_X), \eta, \lambda_X)$ and $(\Theta_X, \mathscr{B}(\Theta_X), \eta, \lambda_X^*)$. From Definition 2.6, we have

$$\lambda_X(\eta_X(A \cup B)) + \lambda_X(\eta_X(A \cap B)) \geq \lambda_X(\eta_X(A)) + \lambda_X(\eta_X(B)). \tag{2.51}$$

We know that

$$\eta_X(A \cap B) \subset \eta_X(A), \quad \eta_X(A \cap B) \subset \eta_X(B), \quad \eta_X(A \cap B) \subset \eta_X(A) \cap \eta_X(B),$$

$$\lambda_X(\eta_X(A \cap B)) \leq \lambda_X(\eta_X(A) \cap \eta_X(B)),$$

and from (2.51), we obtain

$$\lambda_X(\eta_X(A) \cup \eta_X(B)) = \lambda_X(\eta_X(A)) + \lambda_X(\eta_X(B)) - \lambda_X(\eta_X(A) \cap \eta_X(B))$$

$$\leq \lambda_X(\eta_X(A \cup B)) + \lambda_X(\eta_X(A \cap B)) - \lambda_X(\eta_X(A) \cap \eta_X(B)) \leq \lambda_X(\eta_X(A \cup B)).$$

We can prove the second inequality of (2.50) analogously.

Sufficiency. Suppose that for for every pair of probability measures $\lambda_X \in L_X^g$, $\lambda_X^* \in L_X^{g^*}$, inequality (2.50) is satisfied. Let $(\Theta^*, \mathscr{B}(\Theta^*), \eta^*, \lambda^*)$ be any MSR of the monotone measure g^*. If $\lambda_X^{*0} \in L_X^{g^*}$ is equivalent to λ^*, then we have $\forall A, B \in \mathscr{B}(X)$,

$$\begin{aligned} g^*(A \cup B) &= \lambda^*(\eta^*(A \cup B)) = \lambda^*(\tau^{*-1}(\eta_X^{*0}(A \cup B))) \\ &= \lambda_X^{*0}(\eta_X^{*0}(A \cup B)) \leq \lambda_X^{*0}(\eta_X^{*0}(A) \cup \eta_X^{*0}(B)) \\ &= \lambda_X^{*0}(\eta_X^{*0}(A)) + \lambda_X^{*0}(*\eta_X^{*0}(B)) - \lambda_X^{*0}(\eta_X^{*0}(A) \cap \eta_X^{*0}(B)) \\ &\leq \lambda_X^{*0}(\eta_X^{*0}(A)) + \lambda_X^{*0}(\eta_X^{*0}(B)) - \lambda_X^{*0}(\eta_X^{*0}(A \cap B)) \end{aligned}$$

$$= \lambda_X^{*0}(\tau^*(\eta^*(A))) + \lambda_X^{*0}(\tau^*(\eta^*(A \cap B)))$$

$$= \lambda^*(\eta^*(A)) + \lambda^*(\eta^*(B)) - \lambda^*(\eta^*(A \cap B))$$

$$= g^*(A) + g^*(B) - g^*(A \cap B),$$

i.e., g^* is Choquet upper capacity of order two. We can prove the first inequality of Definition 2.6 analogously. □

Definition 2.13. NRCs L_X^g and $L_X^{g^*}$ of a pair dual monotone measures g, g^* are called dual classes of inequivalent probability representations.

We note that a monotone measure $g : \mathscr{B}(X) \to [0,1]$ is a probability measure if and only if $L_X^g \cap L_X^{g^*} \neq \varnothing$. The proof is trivial.

Let F be any fuzzy subset on X, and let $F(\cdot)$ be its compatibility function: $F(\cdot) : X \to [0,1]$. If $\sigma \in S_n$ is a permutation such that

$$F(x_{\sigma(1)}) \leq F(x_{\sigma(2)}) \leq \cdots \leq F(x_{\sigma(n)})$$

and $K_i \equiv \{x_{\sigma(i)}, x_{\sigma(i+1)}, \ldots, x_{\sigma(n)}\}$, $i = 1, 2, \ldots, n$, are nested subsets of X, then monotone expectation may be represented as

$$E_g(F) = \sum_{i=1}^{n} F(x_{\sigma(i)})\{g(K_i) - g(K_{i+1})\},$$

where $g(K_{n+1}) \equiv 0$. Choquet's integral has been studied in [8, 16, and others].

Proposition 2.16 ([8]). *If $\{P_\sigma\}_{\sigma \in S_n}$ is an APC of the monotone measure g, then $\exists \sigma_0 \in S_n$,*

$$E_g(F(\cdot)) = E_{P_{\sigma_0}}(F(\cdot)) = \int_X F\,dP_{\sigma_0}. \tag{2.52}$$

A monotone expectation is represented as a mathematical expectation of $F(\cdot)$ with respect to the probability measure P_{σ_0}. We have a somewhat similar proposition for MSR:

Proposition 2.17 ([93]). *If $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$ is any MSR of the monotone measure g, then there exists a nonnegative function $\widehat{F}(\cdot)$ on Θ such that*

$$E_g(F(\cdot)) = (c)\int_X F(\cdot)\,dg = \int_\Theta \widehat{F}(\cdot)\,d\lambda = E_\lambda(\widehat{F}(\cdot)), \tag{2.53}$$

where $\forall \theta \in \Theta$:

$$\widehat{F}(\theta) = \sup\{\alpha \mid \theta \in \eta(\{x \mid F(x) \geq \alpha,\ 0 \leq \alpha \leq 1\})\}. \tag{2.54}$$

Choquet's integral is represented as a Lebesgue integral on Θ with respect to the probability measure λ.

Analogously to Proposition 2.9, we have the following results.

Proposition 2.18. *Let F_1 and F_2 be any fuzzy subsets on X with compatibility functions $F_1(\cdot)$, $F_2(\cdot)$. A pair of dual monotone measures (g, g^*) are Choquets's lower and upper capacities of order two respectively if and only if*

$$\begin{aligned} E_g(F_1(\cdot)+F_2(\cdot)) &\geq E_g(F_1(\cdot))+E_g(F_2(\cdot)), \\ E_{g^*}(F_1(\cdot)+F_2(\cdot)) &\leq E_{g^*}(F_1(\cdot))+E_{g^*}(F_2(\cdot)). \end{aligned} \tag{2.55}$$

Proof. Necessity. Let g, g^* be Choquet dual capacities of order two such that Proposition 2.15 and (2.50) are true. Using the properties of the supremum function and mathematical expectation, we have

$$\begin{aligned} E_g(F_1(\cdot)+F_2(\cdot)) &= E_\lambda\left(\sup\{\alpha \mid \theta \in \eta(\{x \mid F_1(x)+F_2(x) \geq \alpha\}),\ 0 \leq \alpha \leq 1\}\right) \\ &\geq E_\lambda\left(\sup\left\{\alpha \mid \theta \in \eta\left(\left\{x \mid F_1(x) \geq \frac{\alpha}{2}\right\} \cup \left\{x \mid F_2(x) \geq \frac{\alpha}{2}\right\}\right)\right\}\right) \\ &\geq E_\lambda\left(\sup\left\{\frac{\alpha}{2} \mid \theta \in \eta\left(\left\{x \mid F_1(x) \geq \frac{\alpha}{2}\right\}\right)\right\}\right. \\ &\quad \left.+\left\{\frac{\alpha}{2} \mid \theta \in \eta\left(\left\{x \mid F_2(x) \geq \frac{\alpha}{2}\right\}\right)\right\}\right) \\ &\geq E_\lambda\left(\sup\left\{\alpha' \mid \theta \in \eta\left(\left\{x \mid F_1(x) \geq \frac{\alpha}{2}\right\}\right)\right\}\right. \\ &\quad \left.+\left\{\alpha' \mid \theta \in \eta\left(\left\{x \mid F_2(x) \geq \frac{\alpha}{2}\right\}\right)\right\}\right) = E_g(F_1(\cdot))+E_g(F_2(\cdot)). \end{aligned}$$

We may obtain the second inequality of (2.55) similarly.

Sufficiency. Let the inequalities (2.55) be satisfied $\forall A, B \in \mathscr{B}(X)$. If $F_1(\cdot) \equiv I_A$, $F_2(\cdot) \equiv I_B$, where I_A and I_B are the indicator functions of sets A and B respectively, then we have

$$E_g(I_A+I_B) \geq E_g(I_A)+E_g(I_B).$$

It is easily proved that

$$E_g(I_{A\cup B}+I_{A\cap B}) = E_g(I_{A\cup B})+E_g(I_{A\cap B}).$$

Note that $\forall x \in X$, $I_A(x)+I_B(x) = I_{A\cup B}+I_{A\cap B}$, and from the property of a monotone expectation, we have $E_g(I_A+I_B) = E_g(I_{A\cup B}+I_{A\cap B})$.

We also have

$$\begin{aligned} g(A)+g(B) = E_g(I_A)+E_g(I_B) &\leq E_g(I_A+I_B) = E_g(I_{A\cup B}+I_{A\cap B}) \\ &= E_g(I_{A\cup B})+E_g(I_{A\cap B}) = g(A\cup B)+g(A\cap B), \end{aligned}$$

i.e., g is Choquet lower capacity of order two. Therefore, g^* will be an upper capacity of order two, which follows from the second inequality of (2.55). □

2.2.2 Distance Between Monotone Measures in MSR

Distances in the set of monotone measures are defined through distances between their APCs [9]:

$$D(g_1, g_2) = D(\{P^1_\sigma(\cdot)\}_{\sigma \in S_n}; \{P^2_\sigma(\cdot)\}_{\sigma \in S_n}), \tag{2.56}$$

where $g_1, g_2 \in [0,1]^{\mathscr{B}(X)}$ are monotone measures, $\{P^1_\sigma(\cdot)\}_{\sigma \in S_n}$, and $\{P^2_\sigma(\cdot)\}_{\sigma \in S_n}$ are respectively APCs of g_1 and g_2. For example the distance D_2 is given by

$$D_2^2(g_1, g_2) = \sum_{\sigma \in S_n} \sum_{i=1}^{n} \left(P^{(1)}_\sigma(x_{\sigma(i)}) - P^{(2)}_\sigma(x_{\sigma(i)}) \right)^2. \tag{2.57}$$

Practically, the distance between monotone measures g_1 and g_2 is reduced to known distances between probability measures [9].

Let $\{(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X)_{\lambda_X \in L_X^{g_i}}\}$, $i = 1, 2$, be NRCs of monotone measures g_1 and g_2, respectively. We introduce a new distance between the monotone measures g_1, g_2:

Definition 2.14. A distance between monotone measures is the distance between classes $L_X^{g_1}$ and $L_X^{g_2}$ of inequivalent probability representations:

$$D^2(g_1, g_2) = D^2(L_X^{g_1}, L_X^{g_2}) = \inf_{\lambda_X^{(1)} \in L_X^{g_1}, \lambda_X^{(2)} \in L_X^{g_2}} D^2(\lambda_X^{(1)}, \lambda_X^{(2)}). \tag{2.58}$$

In this subsection, our problem is to parameterize the distance D in order to calculate the infimum in (2.58). Let Θ be a “specific” set [93],

$$\Theta = \left\{ \lambda_{\overline{0}}, \lambda_{\overline{1}}, \ldots, \lambda_{\overline{n}}, \lambda_{\overline{12}}, \ldots, \lambda_{\overline{n-1,n}}, \lambda_{\overline{123}}, \ldots, \lambda_{\overline{n-2,n-1,n}}, \ldots, \lambda_{\overline{12\ldots n}} \right\}, \tag{2.59}$$

and let η be constructed as

$$\eta(\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}) = \left\{ \overline{0}, \overline{i}_1, \overline{i}_2, \ldots, \overline{i}_k, \overline{i_1 i_2}, \ldots, \overline{i_{k-1}, i_k}, \ldots, \overline{i_1, \ldots, i_{n-1}, i_n} \right\},$$

where the operation $\overline{ab}$ on a and b is a “concatenation” of numbers a and b. Then $\forall A = \{x_{i_1}, \ldots, x_{j_k}\} \in \mathscr{B}(X)$,

$$g(A) = \sum_{\theta \in \eta(A)} \lambda_\theta = \lambda_{\overline{0}} + \lambda_{\overline{i}_1} + \lambda_{\overline{i}_2} + \cdots + \lambda_{\overline{i}_k} + \lambda_{\overline{i_1 i_2}} + \cdots + \lambda_{\overline{i_{k-1} i_k}} + \lambda_{\overline{i_1 i_2 \ldots i_k}}. \tag{2.60}$$

We may represent (2.60) as a sum of “parts” of elements $x_{i_j} \in X$ whose indices are constructed by all subsets $B \subset A$ such that $x_{i_j} \in B$. If $g : \mathscr{B}(X) \to [0,1]$ is a known monotone measure, then the scheme of finding parameters λ is as follows:

$$
\begin{aligned}
& g_1 \equiv g(\{x_1\}) = \lambda_{\overline{0}} + \lambda_{\overline{1}}, \\
& \dots\dots\dots\dots\dots \\
& g_n \equiv g(\{x_n\}) = \lambda_{\overline{0}} + \lambda_{\overline{n}}, \\
& g_{12} \equiv g(\{x_1, x_2\}) = \lambda_{\overline{0}} + \lambda_{\overline{1}} + \lambda_{\overline{2}} + \lambda_{\overline{12}}, \\
& \dots\dots\dots\dots\dots \\
& g_{n-1,n} \equiv g(\{x_{n-1}, x_n\}) = \lambda_{\overline{0}} + \lambda_{\overline{n-1}} + \lambda_{\overline{n}} + \lambda_{\overline{n-1,n}}, \\
& g_{123} \equiv g(\{x_1, x_2, x_3\}) = \lambda_{\overline{0}} + \lambda_{\overline{1}} + \lambda_{\overline{2}} + \lambda_{\overline{3}} + \lambda_{\overline{12}} + \lambda_{\overline{13}} + \lambda_{\overline{23}} + \lambda_{\overline{123}}, \\
& \dots\dots\dots\dots\dots \\
& 1 \equiv g(\{x_1, x_2, \dots, x_n\}) = \lambda_{\overline{0}} + \lambda_{\overline{1}} + \cdots + \lambda_{\overline{n}} + \lambda_{\overline{12}} + \cdots + \lambda_{\overline{n-1,n}} \\
& \qquad\qquad + \lambda_{\overline{123}} + \cdots + \lambda_{\overline{12\dots n}}.
\end{aligned} \tag{2.61}
$$

We have $2^n - 1$ equations in 2^n unknown parameters λ. One parameter is free, and for convenience it is chosen as λ_0. Then

$$
\begin{aligned}
& \lambda_1 = g_{\overline{1}} - \lambda_{\overline{0}}, \\
& \lambda_2 = g_{\overline{2}} - \lambda_{\overline{0}}, \\
& \dots\dots\dots\dots\dots \\
& \lambda_n = g_{\overline{n}} - \lambda_{\overline{0}}, \\
& \lambda_{\overline{12}} = g_{\overline{12}} - \lambda_{\overline{1}} - \lambda_{\overline{2}} - \lambda_{\overline{0}}, \\
& \dots\dots\dots\dots\dots \\
& \lambda_{\overline{n-1,n}} = g_{\overline{n-1,n}} - \lambda_{\overline{n-1}} - \lambda_{\overline{n}} - \lambda_{\overline{0}}, \\
& \lambda_{\overline{123}} = g_{\overline{123}} - \lambda_{\overline{1}} - \lambda_{\overline{2}} - \lambda_{\overline{3}} - \lambda_{\overline{12}} - \lambda_{\overline{13}} - \lambda_{\overline{23}} - \lambda_{\overline{123}}, \\
& \dots\dots\dots\dots\dots \\
& \lambda_{\overline{12\dots n}} = g_{\overline{12\dots n}} - \lambda_{\overline{1}} - \lambda_{\overline{2}} + \lambda_{\overline{3}} - \cdots - \lambda_{\overline{23\dots n}} - \lambda_{\overline{0}}.
\end{aligned} \tag{2.62}
$$

It is clear that $\forall A \in \mathscr{B}(X)$,

$$
g(A) = \lambda(\eta(A)) \equiv f(\lambda_{\overline{0}}), \tag{2.63}
$$

and f is a linear function of $\lambda_{\overline{0}}$. Let $0 \le M_0^- \le \lambda_{\overline{0}} \le M_0^+ \le 1$. Let g_1 and g_2 be monotone measures on $\mathscr{B}(X)$. We know that $\forall A \in \mathscr{B}(X)$, $\lambda_X(\eta_X(A)) = \lambda(\tau^{-1}(A))$, and the value of λ_X is also a linear function of the parameter λ_0. Let $\lambda_X(\eta_X(A)) = \lambda_0 + \widehat{\lambda}(A)$, where $\widehat{\lambda}(A)$ is known as a certain expression of parameters $\lambda_{\overline{1}}, \dots, \lambda_{\overline{12\dots n}}$

and numbers that may be calculated by (2.60). Let $D \equiv D_2$ between probability measures. By transformation (2.58), we obtain

$$\begin{aligned} D_2^2(g_1, g_2) &= \inf_{\lambda_X^{(1)} \in L_X^{g_1}, \lambda_X^{(2)} \in L_X^{g_2}} D_2^2(\lambda_X^{(1)}, \lambda_X^{(2)}) \\ &\stackrel{\text{def}}{=} \inf_{\lambda_X^{(1)} \in L_X^{g_1}, \lambda_X^{(2)} \in L_X^{g_2}} \sum_{i=1}^{n} (\lambda_X^{(1)}(x_i) - \lambda_X^{(2)}(x_i))^2. \end{aligned} \tag{2.64}$$

Then

$$\begin{aligned} D_2^2(g_1, g_2) &= \inf_{\substack{M_1^- \le \lambda_0^{(1)} \le M_1^+ \\ M_2^- \le \lambda_0^{(1)} \le M_2^+}} D_2^2(\lambda_0^{(1)} + \lambda_1(\cdot), \lambda_0^{(2)} + \lambda_2(\cdot)) \\ &= \inf_{\substack{M_1^- \le \lambda_0^{(1)} \le M_1^+ \\ M_2^- \le \lambda_0^{(2)} \le M_2^+}} \sum_{i=1}^{n} \Big((\lambda_0^{(1)} - \lambda_0^{(2)}) + (\lambda_1(x_i) - \lambda_2(x_i)) \Big)^2 \\ &= \inf_{\substack{M_1^- \le \lambda_0^{(1)} \le M_1^+ \\ M_2^- \le \lambda_0^{(2)} \le M_2^+}} \Big\{ n(\lambda_0^{(1)} - \lambda_0^{(2)}) + 2(\lambda_0^{(1)} - \lambda_0^{(2)}) \sum_{i=1}^{n} (\widehat{\lambda}_1(x_i) - \widehat{\lambda}_2(x_i)) \\ &\qquad + \sum_{i=1}^{n} ((\widehat{\lambda}_1(x_i) - \widehat{\lambda}_2(x_i))^2 \Big\}. \end{aligned}$$

Define

$$\lambda_0^{(1)} - \lambda_0^{(2)} \equiv \lambda_0, \quad \sum_{i=1}^{n} (\widehat{\lambda}_1(x_i) - \widehat{\lambda}_2(x_i)) \equiv b, \quad \sum_{i=1}^{n} (\widehat{\lambda}_1(x_i) - \widehat{\lambda}_2(x_i))^2 \equiv c.$$

Then

$$\begin{aligned} D_2^2(g_1, g_2) &= \inf_{M_1^- - M_2^+ \le \lambda_0 \le M_1^+ - M_2^-} \left\{ n\lambda_0^2 + 2b\lambda_0 + c \right\} \\ &= \begin{cases} \min \left\{ n(M^-)^2 + 2bM^- + c; n(M^+)^2 + 2bM^+ + c \right\}, & \text{if } -\frac{b}{n} \notin [M^-; M^+], \\ \dfrac{nc - b^2}{n}, & \text{if } -\frac{b}{n} \in [M^-; M^+], \end{cases} \end{aligned}$$

where $M^- \equiv M_1^- - M_2^+$, $M^+ = M_1^+ - M_2^-$.

The following proposition on the correctness of the definition of the distance between monotone measures (as in [9]) is easily proved:

Proposition 2.19. *If g_1 and g_2 are any monotone measures on $\mathscr{B}(X)$, then*

$$D_1(g_1, g_2) = D_2(g_1^*, g_2^*). \tag{2.65}$$

So we may consider monotone measures and their dual monotone measures to contain the same information but codified in different ways.

2.2.3 Connection Between CBR and MSR

It is clear that $\forall \sigma \in S_n$, $i = 1, 2, \ldots, n$,

$$P_\sigma(x_{\sigma(i)}) = \lambda(\eta(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\})) - \lambda(\eta(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\})), \tag{2.66}$$

where $\lambda(\eta(\{x_{\sigma(0)}\})) \equiv 0$, i.e., if MSR $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$ of the monotone measure g is known, then from (2.66) we obtain APC $\{P_\sigma\}_{\sigma \in S_n}$of CBR. In contrast, if the APC $\{P_\sigma\}_{\sigma \in S_n}$ of the CBR of the monotone measure g is known, then we have the following.

Proposition 2.20. *If $\{P_\sigma\}_{\sigma \in S_n}$ is APC of the monotone measure g, then there exists an MSR-$(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$ that is induced by CBR.*

Proof. We construct the set

$$\Theta = \{P_\sigma(B) > 0 \mid \sigma \in S_n,\ B \in \mathscr{B}(X)\} \tag{2.67}$$

and consider a probability measure λ on $\mathscr{B}(\Theta) : \forall P_\sigma(B) > 0$,

$$\lambda(\{P_\sigma(B)\}) = P_\sigma(B) - \max_{\Theta_B^\sigma} P_\beta(G), \tag{2.68}$$

where

$$\Theta_B^\sigma = \{P_\beta(G) \mid \beta \in S_n,\ G \in \mathscr{B}(X),\ P_\beta(G) < P_\sigma(B)\}.$$

It is clear that λ is a probability measure on $\mathscr{B}(X)$. Then $\forall A \in \mathscr{B}(X)$,

$$\lambda(\eta(A)) = \sum_{P_\sigma(B) \in \eta(A)} \lambda(\{P_\sigma(B)\}), \tag{2.69}$$

but $\eta(A)$ is constructed as

$$\eta(A) = \{P_\sigma(B) > 0 \mid \sigma \in S_n,\ B \in \mathscr{B}(X),\ P_\sigma(B) \le g(A)\}. \tag{2.70}$$

With regard to (2.68), (2.69) may be simplified:

$$\lambda(\eta(A)) = \sum_{\substack{P_\sigma(B) \le P_\sigma(A) \\ \sigma \in S_n}} \left\{ P_\sigma(B) - \max_{\Theta_B^\sigma} P_\beta(G) \right\} = P_{\sigma_A}(A) \equiv g(A). \tag{2.71}$$

The proposition is proved. $\square$

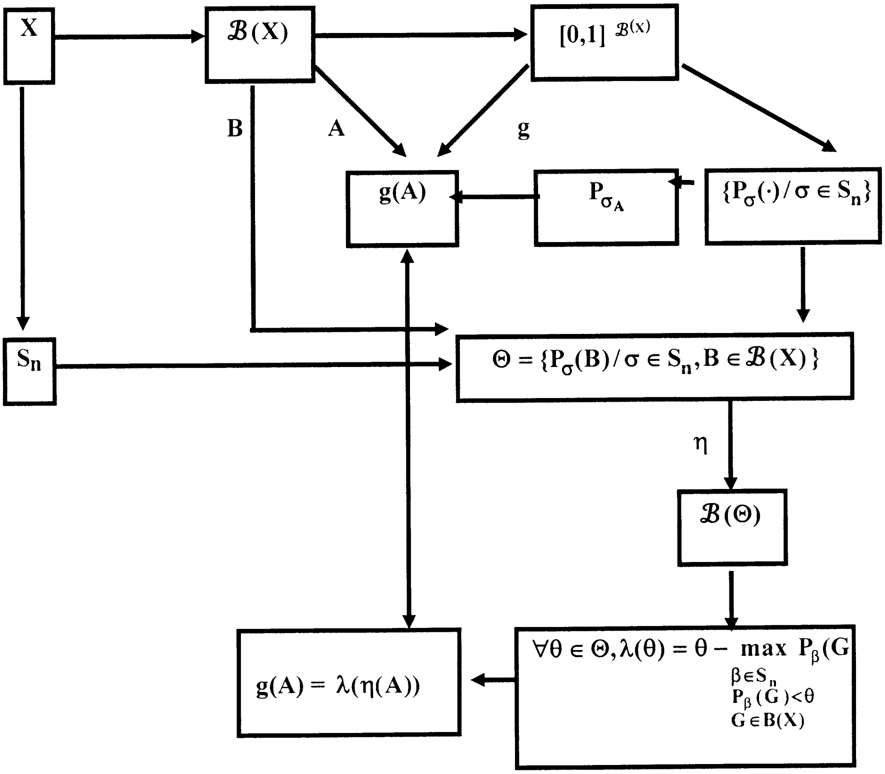

Fig. 2.1 The schematic connection between CBR and MRS

So the constructed MSR is called the representation induced by APC. The schematic connection between CBR and MSR is presented in Fig. 2.1. From Fig. 2.2 we have $\forall A \in \mathscr{B}(X)$,

$$\begin{aligned} \xi \circ \rho(A) &= P_{\sigma_A}(A) = g(A), \\ \lambda \circ \eta(A) &= \lambda(\eta(A)) = g(A), \\ \lambda \circ \mu \circ \rho(A) &= \lambda(\mu(\rho(A))) = g(A). \end{aligned} \tag{2.72}$$

2.2.4 An Example of Connections

For a clearer representation of the facts obtained, above we shall consider an example in which $X = \{x_1, x_2\}$. The semifilters are

$$S_1 = \{\{x_1\}, X\}, \quad S_2 = \{\{x_2\}, X\}, \quad S_3 = \{X\}.$$

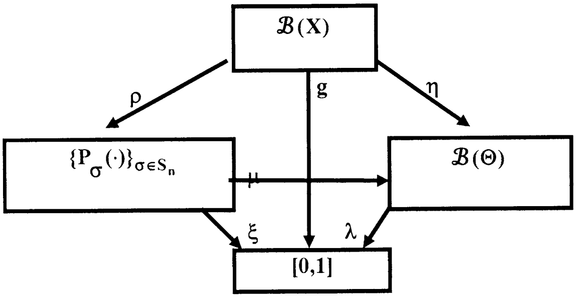

Fig. 2.2 The compositional connection between CBR and MRS

Then

$$\Theta_X = \{S_1, S_2, S_3\}, \quad \mathscr{B}(\Theta_X) = \{\varnothing, S_1, S_2, S_3, \Theta_X\}, \quad \forall A \in \mathscr{B}(X):$$

$$\begin{aligned}
\eta_X(A) &= \{S \in \Theta_X \mid A \in S\};\\
\eta(\{x_1\}) &= \{S_1\} = \{\{x_1\}, X\},\\
\eta(\{x_2\}) &= \{S_2\} = \{\{x_2\}, X\},\\
\eta(X) &= \{S_1, S_2, S_3\} = \Theta_X,\\
\eta(\varnothing) &= \varnothing.
\end{aligned}$$

And η_X is a 0–1 order-preserving homomorphism.

Let $g_1 = g(\{x_1\})$, $g_2 = g(\{x_2\})$, $g_1^* = g^*(\{x_1\})$, $g_2^* = g^*(\{x_2\})$; let $(\Theta_X, \mathscr{B}(\Theta_X), \eta_X, \lambda_X)$ be MSR of g. Then

$$\begin{aligned}
g_1 &= \lambda_X(\eta_X(\{x_1\})) = \lambda_X(\{S_1\}) = 1 - g_2^*,\\
g_2 &= \lambda_X(\eta_X(\{x_2\})) = \lambda_X(\{S_2\}) = 1 - g_1^*,\\
\lambda_1(\{S_1 \cup S_2\}) &= \lambda_X(\{X\}) - 1 - g_1 - g_2,\\
\lambda(\Theta_X) &\equiv 1.
\end{aligned}$$

This representation is schematically shown in Fig. 2.3. From (2.59), we have the MSR $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$, where $\Theta = \{0, 1, 2, 12\}$ and

$$\begin{aligned}
\eta(\{x_1\}) &= \{0,1\} \Rightarrow g_1 = g(\{x_1\}) = \lambda_0 + \lambda_1,\\
\eta(\{x_2\}) &= \{0,2\} \Rightarrow g_2 = g(\{x_2\}) = \lambda_0 + \lambda_2,\\
\eta(\{X\}) &= \{0,1,2,12\} \Rightarrow 1 = g(X) = \lambda_0 + \lambda_1 + \lambda_2 + \lambda_{12}.
\end{aligned}$$

We obtain the system of equations

$$\begin{cases}
g_1 = \lambda_0 + \lambda_1,\\
g_2 = \lambda_0 + \lambda_2,\\
1 = \lambda_0 + \lambda_1 + \lambda_2 + \lambda_{12}.
\end{cases}$$

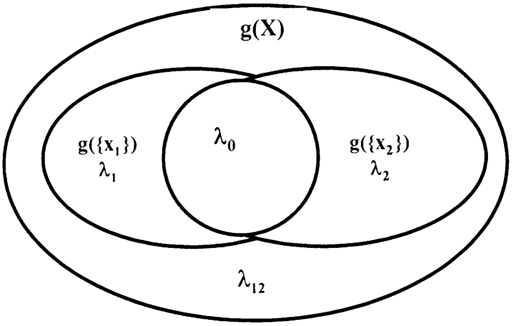

Fig. 2.3 The schematic connection between the monotone measure g and parameters λ

Table 2.1 APC of the monotone measure g

σ	$x_{\sigma(1)}$	$x_{\sigma(2)}$
$(1,2)$	$P_1^1 = P_1(x_{\sigma(1)})$	$P_1^2 = P_1(x_{\sigma(2)})$
$(2,1)$	$P_2^1 = P_2(x_{\sigma(2)})$	$P_2^2 = P_2(x_{\sigma(2)})$

We have three equations in four unknowns. For the calculation of $D_2^2(g_1, g_2)$, where g_1, g_2 are two measures on $\mathscr{B}(X)$, we get

$$D_2^2(g_1,g_2) = \inf_{-1\le\lambda_0\le 1} \{4\lambda_0^2 - 4(\widetilde{g}_1+\widetilde{g}_2)\lambda_0 + [\widetilde{g}_1^2+\widetilde{g}_2^2+(\widetilde{g}_1^2+\widetilde{g}_2^2)^2]\} = \widetilde{g}_1^2+\widetilde{g}_2^2,$$

where $\lambda_{0(\min)} = \frac{\widetilde{g}_1+\widetilde{g}_2}{2}$, $\widetilde{g}_1 \equiv g_1^{(1)} - g_1^{(2)}$, $\widetilde{g}_2 \equiv g_{21}^{(1)} - g_1^{(2)}$. It is clear that $\widetilde{g}_1 = \widetilde{g}_1^*$, $\widetilde{g}_2 = \widetilde{g}_2^*$, and $D_2(g_1,g_2) = D_2(g_1^*,g_2^*)$. Let $\{P_\sigma(\cdot)\}_{\sigma\in S_2}$ be APC of g: $P_\sigma(x_{\sigma(1)}) = g(\{x_{\sigma(1)}), P_\sigma(x_{\sigma(2)}) = 1 - g(\{x_{\sigma(2)})$ (see Table 2.1). We have similarly APC of g^*: $\forall A \in \mathscr{B}(X)$,

$$g(A) = \lambda(\eta(A)) = P_{\sigma_A}(A) \Leftrightarrow \begin{cases} g_1 = \lambda_0 + \lambda_1 = P_1^1, \\ g_2 = \lambda_0 + \lambda_2 = P_2^1, \\ \lambda_0+\lambda_1+\lambda_2+\lambda_{12} = P_1^1+P_1^2 = P_2^1+P_2^2 = 1, \end{cases}$$

and from (2.67), $\Theta = \{P_1^1, P_1^2, P_2^1, P_2^2\}$. Let $P_1^1 < P_1^2 < P_2^1 < P_2^2 < 1$. Then $\lambda(P_1^1) = P_1^1$, $\lambda(P_2^1) = P_2^1 - P_1^1$, $\lambda(P_1^2) = P_2^2 - P_1^2$. We construct the MSR $(\Theta, \mathscr{B}(\Theta), \eta, \lambda)$: $\forall\theta \in \Theta$,

$$\tau(\theta) = \{A \in \mathscr{B}(X) \mid \theta \in \eta(A)\}$$

and

$$\tau(P_1^1) = \{\{x_1\}, X\} = S_1,$$

$$\tau(P_2^1) = \{\{x_2\}, X\} = S_2,$$

$$\tau(P_1^1) = \{x_1\} \cup \{x_2\} = X,$$

$$\tau(P_2^2) = S_1 \cup S_2.$$

Then $\forall E \in \mathscr{B}(\Theta_X)$, we have $\lambda_X(E) = \lambda(\tau^{-1}(E))$ and

$$\begin{aligned}
\lambda_X(S_1) &= \lambda(P_1^1) = P_1^1, \\
\lambda_X(S_2) &= \lambda(P_2^1) = P_2^1, \\
\lambda_X(X) &= \lambda(P_1^2) = 1 - P_1^1 = P_1^2, \\
\lambda_X(S_1 \cup S_2) &= \lambda(P_2^2) = 1 - P_2^1 = P_1^2, \\
\lambda(\Theta_X) &\equiv 1.
\end{aligned}$$

Conclusion: Two probability representations of monotone measures on finite sets CBR and MSR have been considered. In order to study properties of a concrete class of monotone measures (in our case, Choquet capacities of order two), CBR is "more convenient" as a tool than MSR. But some theorems for Choquet second-order capacities in CBR can be translated for MSR.

In MSR, the notion of a distance between monotone measures is introduced as the distance between inequivalent representation classes (NPC). After a parameterization procedure, it takes an "almost" clear form, which is more convenient for calculations than is possible in CBR.

2.3 Monotone Measure Restoration Problem Based on the CBR

This section deals with problems of monotone measure restoration (or identification) [42, 43, 72, 96, 128, 142, 154, 156, 159, and others] from insufficient data on a finite set. The proposed approach is constructed in the class of second-order Choquet capacities when the "fuzzy weights" of singletons are known. This essentially concerns certain frequency distributions, where the nature of additivity is doubtful because of the fuzzy nature of the data distribution. This is an indispensable condition for the introduction of a monotone measure, but it is an insufficient one for its construction.

Measures of specificity, indices of uncertainty, and estimators of approximations are calculated. Some approximation properties are proved.

2.3.1 Introduction

It is well known that if experimental data are represented by intervals and form a so-called consonant or combined (consonant–dissonant) body of evidence [27, 68, 70, 83, 110, and others], then their distribution is fuzzy and characterized by overlaps. If experimental data are obtained and described using simultaneously

Table 2.2 The insufficient expert frequency distribution of some disease Y with four symptoms x_1, x_2, x_3, x_4 in terms of the monotone measure g

$A \subseteq X = \{x_1, x_2, x_3, x_4\}$	$g(\{\cdot\})$
$\{x_1\}$	0.2
$\{x_2\}$	0.3 subj.[a]
$\{x_3\}$	0.4
$\{x_4\}$	0.2 subj.[a]
$\{x_1, x_2\}$	?
$\{x_1, x_3\}$	?
$\{x_1, x_4\}$	?
$\{x_2, x_3\}$	?
$\{x_2, x_4\}$	?
$\{x_3, x_4\}$	?
$\{x_1, x_2, x_4\}$	?
$\{x_2, x_3, x_4\}$	?
$\{x_1, x_2, x_3\}$	?
$\{x_1, x_3, x_4\}$	?
$\{x_1, x_2, x_3, x_4\}$	1

[a]The abbreviation “subj. ” means that data are appointed by the expert

objective estimations and expert data, then the nature of the data becomes combined, i.e., along with probabilistic–statistical uncertainty there also exists possibility uncertainty. Clearly, in that case a satisfactory result can be obtained only if one resorts to probabilistic–possibilistic analysis, in which an important role belongs to fuzzy statistics successfully used in decision-making supporting systems.

Monotone (fuzzy) measures were first used by Sugeno [141], whose integral A. Kandel called fuzzy statistics [57]. An obligatory attribute of fuzzy statistics is a monotone measure with respect to which they are calculated. The construction of a monotone measure is the most important task in fuzzy analysis [43, 72, 96, 128, 142, 154, 156, 159].

In this section we consider the following problem: to restore monotone measure using expert-objective insufficient data (Sect. 2.3.3). The most successful classification of the methods of monotone measure construction is given in [72], where the main directions of solution of this problem are outlined. Here we are interested in restoring monotone measure on the basis of expert knowledge.

In real situations, in describing subjective information, expert data are frequently connected only with “one-element” factors or almost only with them, since the reception of expert data is connected with multifactor measurements, which for understandable reasons practically do not exist. For example, for some disease Y that is characterized by four factors x_1, x_2, x_3, x_4, an examining physician (expert) can represent the frequency distribution table resulting from these factors (see Table 2.2, where “subj.” denotes weights assigned by the expert), where some “weights” are subjectively assigned on the basis of experience, knowledge, intuition (due to his activity), while there exist no “fuzzy weights” for symptoms of two or more elements (by “fuzzy weights” we understand here values of the corresponding monotone measure on individual elements). Certainly, in the case of another expert, the distribution of frequencies may be different. There exists a possibility that the sum of frequencies of all one-element factors is not equal to one. This fact serves as a criterion of the nonprobabilistic nature of the distribution. It is natural that in the case of such a representation of information, one should use nonadditive monotone measures.

Our investigation here involves the case in which only one-element factors have known "fuzzy weights." For combined variants, because of the nonadditivity of a monotone measure, one cannot calculate "fuzzy weights" as a simple sum using the "fuzzy weights" of the one-element factors that these variants contain. The monotone measure remains partly unknown. Therefore, the problem of monotone measure restoration in a certain class of monotone measures and in the conditions of the existing uncertainty is an important one.

We propose a method of restoring the unknown monotone measure in the class of λ-additive measures in the best approximation to probability measures on the condition that only the "fuzzy weights" of one-element factors are known.

In Sect. 2.3.2 we recall the basic notions and definitions needed for our consideration. In Sect. 2.3.3, monotone measures of optimal approximations of zeroth, first, and second approaches are constructed. The corresponding index of specificity as well as approximation errors are calculated. Correctness theorems are proved for dual measures of optimal approximation.

2.3.2 Preliminary Concepts

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite reference set, $\mathscr{B}(X)$ the full power set of X, g the monotone measure on $\mathscr{B}(X)$ in the sense of Sugeno, and $(X, \mathscr{B}(X), g)$ the monotone measure space [141]. We repeat some definitions from previous sections:

1°. A monotone measure $g_\lambda \in [0,1]^{\mathscr{B}(X)}$ $(\lambda > -1)$ is a λ-additive monotone measure if for every $A, B \in \mathscr{B}(X)$, $A \cap B = \varnothing$,

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) \cdot g_\lambda(B). \tag{2.73}$$

It is easy to verify that for every $A \in \mathscr{B}(X)$,

$$g_\lambda(A) = \frac{1}{\lambda}\left\{\prod_{x_i \in A}(1 + \lambda \widehat{g}_i) - 1\right\}, \tag{2.74}$$

where $0 < \widehat{g}_i \equiv g(\{x_i\}) < 1$; $\lambda > -1$ is a parameter with the following normalization condition:

$$\frac{1}{\lambda}\left\{\prod_{x_i \in X}(1 + \lambda \widehat{g}_i) - 1\right\} = 1. \tag{2.74$'$}$$

Note that g_0 is a probability measure if $\sum\limits_{x_i \in X} \widehat{g}_i = 1$.

2°. The dual monotone measures $g, g^* \in [0,1]^{\mathscr{B}(X)}$ are called respectively the lower and upper Choquet capacities of order two if for every $A, B \in \mathscr{B}(X)$,

$$\begin{aligned} g(A\cap B)+g(A\cup B) &\geq g(A)+g(B), \\ g^*(A\cap B)+g^*(A\cup B) &\leq g^*(A)+g^*(B), \end{aligned} \tag{2.75}$$

where $g^*(A) = 1 - g(\overline{A})$ (duality). Choquet capacities of order two belong to a sufficiently wide class of monotone measures. For example, the λ-additive monotone measure g_λ is a Choquet capacity of order two. It is easy to verify that $g_\lambda^* = g_{-\lambda/(1+\lambda)}$. Let $\{\widehat{g}_i\}$ and $\{\widehat{g}_i^*\}$, $i = 1, 2, \ldots, n$, denote the "fuzzy weights" of singletons for the dual monotone measures g, g^*, respectively.

3°. For each permutation $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n)) \in S_n$ of the finite set $\{1, 2, \ldots, n\}$, the probability functions

$$\begin{aligned} P_\sigma(x_{\sigma(1)}) &= g(\{x_{\sigma(1)}\}), \\ P_\sigma(x_{\sigma(2)}) &= g(\{x_{\sigma(1)}, x_{\sigma(2)}\}) - g(\{x_{\sigma(1)}\}), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ P_\sigma(x_{\sigma(i)}) &= g(\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}) - g(\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\}), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ P_\sigma(x_{\sigma(n)}) &= 1 - g(\{x_{\sigma(1)}, \ldots, x_{\sigma(n-1)}\}) \end{aligned} \tag{2.76}$$

are called the probabilities associated with the monotone measure g, where S_n is the group of permutations of all natural numbers from 1 to n. It is proved [8] that if $g, g^* \in [0,1]^{\mathscr{B}(X)}$ are dual monotone measures, then they have the common class of associated probabilities $\{P_\sigma(\cdot)\}_{\sigma \in S_n}$, and for every $\sigma \in S_n$, we have $P_\sigma(\cdot) = P^*_{\sigma^*}(\cdot)$, where σ^* is the dual permutation of σ ($\sigma(i) = \sigma^*(n-i+1)$, $i = 1, \ldots, n$).

Due to (2.74) and (2.76), we can write the class of associated probabilities for the λ-additive monotone measure g_λ for any $\sigma \in S_n$ as

$$P_\sigma(x_{\sigma(i)}) = g_\lambda(\{x_{\sigma(i)}\}) \prod_{j=1}^{i-1} (1 + \lambda g_\lambda(\{x_{\sigma(j)}\})), \tag{2.77}$$

or more exactly, as

$$P_\sigma(x_i) = g_\lambda(\{x_i\}) \prod_{j=1}^{i(\sigma)-1} (1 + \lambda g_\lambda(\{x_{\sigma(j)}\})), \tag{2.77$'$}$$

where $i = 1, 2, \ldots, p$; $\sigma \in S_n$; $i(\sigma)$ is the location of x_i in the permutation σ (if $i(\sigma) = 1$, then $\prod_{j=1}^{0} \equiv 1$).

4°. Let us introduce the following notation. $\mathfrak{M}(X) \subset [0,1]^{\mathscr{B}(X)}$ is the class of monotone measures on $\mathscr{B}(X)$; $\mathfrak{M}_c(X)$ is the class of Choquet capacities of order two on $\mathscr{B}(X)$; $\mathfrak{M}_\lambda(X)$ is the class of λ-additive monotone measures on $\mathscr{B}(X)$; $\mathfrak{P}(X)$ is the class of probability measures on $\mathscr{B}(X)$. It is clear that $\mathfrak{P}(X) \subset \mathfrak{M}_\lambda(X) \subset \mathfrak{M}_c(X) \subset \mathfrak{M}(X)$. By [8], if $g \in \mathfrak{M}_c(X)$, then for every $A \subseteq X$,

$$g(A) = \min_{\sigma \in S_n} P_\sigma(A), \quad g^*(A) = \max_{\sigma \in S_n} P_\sigma(A).$$

5° [9] Let $\mathbb{T}^m \equiv \{(y_1, y_2, \ldots, y_m) \in \mathbb{R}^m / y_i \geq 0,\, i = 1, 2, \ldots, m\}$. Let f be a function $f : \mathbb{T}^m \to R^+$. We say that f is a function that generates a distance if the following five conditions are satisfied:

(1) $f(y_1, y_2, \ldots, y_m) = 0 \Longleftrightarrow y_1 = y_2 = \cdots = y_m = 0$.
(2) $y_i \leq z_i$, for every $i \Leftrightarrow f(y_1, y_2, \ldots, y_m) \leq f(z_1, z_2, \ldots, z_m)$; f is a nondecreasing monotone function.
(3) $f(y_1 + z_1, y_2 + z_2, \ldots, y_m + z_m) \leq f(y_1, y_2, \ldots, y_m) + f(z_1, z_2, \ldots, z_m)$; f is subadditive.
(4) $f(y, y, \ldots, y) = y$; f is idempotent.
(5) $f(y_{\sigma(1)}, y_{\sigma(2)}, \ldots, y_{\sigma(m)}) = f(y_1, y_2, \ldots, y_m)$, for every $\sigma \in S_m$; f is symmetric.

We rank $n! \equiv m$ permutations of S_n by some criterion in order to number them and thus to represent the class $\{P_\sigma(\cdot)\}_{\sigma \in S_n}$ as an $n!$-dimensional vector $(P_1, P_2, \ldots, P_m)$.

6°. Let d be some distance on $\mathfrak{P}(X)$. It is proved in [9] that a function $D : \mathfrak{M}(X) \times \mathfrak{M}(X) \to R^+$ defined as

$$D_l(g, g') = f(d_l(P_1, P'_1), d_l(P_2, P'_2), \ldots, d_l(P_m, P'_m))$$

is a distance on $\mathfrak{M}(X)$, where l is a parameter of the distance function. The following are examples of such a function f:

$$f_m(y_1, y_2, \ldots, y_m) \equiv \max_{1 \leq i \leq m} \{y_i\},$$

$$f_q(y_1, y_2, \ldots, y_m) \equiv \left(\frac{1}{m} \sum_{i=1}^{m} y_i^q \right)^{1/q}, \quad q \geq 1,$$

while the distance d can be exemplified by

$$d_m(P, P') = \max_{1 \leq i \leq m} |P(x_i) - P'(x_i)|,$$

$$d_q(P, P') = \left(\sum_{i=1}^{m} |P(x_i) - P'(x_i)|^q \right)^{1/q}, \quad q \geq 1,$$

$$d_s(P, P') = \max_{A \subseteq X} |P(x_i) - P'(x_i)|.$$

Let D_2 denote the distance ($q = 2$)

$$D_2(g,g') \equiv \sqrt{\frac{1}{n!}\sum_{\sigma\in S_n}\sum_{i=1}^{m}(P(x_i)-P'(x_i))^2}.$$

The second example of D is

$$D_{mq}(g,g') \equiv \max_{\sigma\in S_n}\left(\sum_{i=1}^{n}|P(x_i)-P'(x_i)|^q\right)^{1/q}.$$

7°. Given $g \in \mathfrak{M}(X)$, the probability measure $P_g \in \mathfrak{P}(X)$ is called the closest one to the monotone measure g if

$$D(g,P_g) = \min_{P\in\mathfrak{P}(X)} D(g,P). \tag{2.78}$$

According to [8], if $P \in \mathfrak{P}(X)$, then the class of associated probabilities contains a single probability distribution $P \equiv P_\sigma$, $\sigma \in S_n$. Hence the problem of minimizing the distance can be reduced to the problem of minimizing the function $D = D_2$ with respect to P:

$$D_2(g,P) = \sqrt{\frac{1}{n!}\sum_{\sigma\in S_n}\sum_{i=1}^{n}(P_\sigma(x_i)-P(x_i))^2} \Rightarrow \min,$$

$P \in \mathfrak{P}(X)$. Applying the well-known classical method of least squares, we obtain the solution

$$P_g(x_i) = \frac{1}{n!}\sum_{\sigma\in S_n} P_\sigma(x_i), \quad i = 1,2,\ldots,n. \tag{2.79}$$

If in (2.79), g is the λ-additive monotone measure g_λ, then with (2.77) taken into account, we have

$$P_{g_\lambda}(x_i) = \frac{\widehat{g}_i}{n!}\sum_{\sigma\in S_n}\prod_{j=1}^{i(\sigma)-1}(1+\lambda\widehat{g}_{\sigma(i)}), \quad i = 1,2,\ldots,n. \tag{2.80}$$

If in (2.80), $i(\sigma) = 1$, then the summand is equal to 1. Here $\widehat{g}_i \equiv g_\lambda(\{x_i\})$. The minimal distance D_2 between the monotone measure g_λ and $\mathfrak{P}(X)$ is

$$\begin{aligned} D_2(g_\lambda,\mathfrak{P}(X)) &= D_2(g_\lambda,P_{g_\lambda}) \\ &= \sqrt{\frac{1}{n!}\sum_{\sigma\in S_n}\sum_{j=1}^{n}\widehat{g}_i^2\left[\prod_{j=1}^{i(\sigma)-1}\{1+\lambda\widehat{g}_{\sigma(j)}\} - \frac{1}{n!}\sum_{\tau\in S_n}\prod_{k=1}^{i(\sigma)-1}\{1+\lambda\widehat{g}_{\tau(k)}\}\right]^2}. \end{aligned} \tag{2.81}$$

If $\lambda = 0$ g_0 is a probability measure (for $\sum_{i=1}^{n} g_i = 1$), then $D_2 \equiv 0$. We call this distance the degree of nonspecificity.

8° [9] For given $g \in \mathfrak{M}(X)$,

$$C(g) = \min\{D(g, Bel_0), D(g^*, Pl_0)\} \tag{2.82}$$

is called the index of specificity, where Bel_0 and Pl_0 are the dual monotone measures of belief and plausibility of the whole ignorance. For any $A \subseteq X$,

$$Bel_0(A) = \begin{cases} 0 \text{ if } A \neq X, \\ 1 \text{ if } A = X, \end{cases} \qquad Pl_0(A) = \begin{cases} 1 \text{ if } A \neq X, \\ 0 \text{ if } A = X. \end{cases}$$

If $C(g_\lambda) \approx 0$, then g_λ is near Bel_0 or Pl_0, and g_λ has no specificity. The class of associated probabilities of Bel_0 is

$$P_\sigma^{Bel_0}(x_{\sigma(i)}) = \begin{cases} 1 \text{ if } i = n, \\ 0 \text{ if } i \neq n, \end{cases} \quad i = 1, 2, \ldots, \quad \sigma \in S_n.$$

Then we obtain

$$P_{Bel_0}(x_i) = \frac{1}{n!} \sum_{\sigma \in S_n} P_\sigma^{Bel_0}(x_{\sigma(i)}) = \frac{1}{n!}(n-1)! = \frac{1}{n},$$

the uniform probability distribution. Therefore

$$C(g_\lambda) = \sqrt{\frac{1}{n!} \sum_{\sigma \in S_n} \sum_{i=1}^{n} \widehat{g}_i^2 \left[\prod_{j=1}^{i(\sigma)-1} \{1 + \lambda \widehat{g}_{\sigma(j)}\} - \frac{1}{n} \right]^2} . \tag{2.83}$$

2.3.3 Problem of Monotone Measure Restoration

We propose a method by which the dual monotone measures (g, g^*) can be restored with the best approach to $\mathfrak{M}_\lambda(X)$ from $\mathfrak{M}(X)$ in terms of the distance D_2 if we impose some additional condition. Let only the "fuzzy weights" of singletons be known (see, for example, Table 2.2):

$$0 < \widehat{g}_i \equiv g(\{x_i\}) < 1, \quad i = 1, 2, \ldots, n. \tag{2.84}$$

Assume that

$$\mathfrak{M}(X, \widehat{g}_1, \widehat{g}_2, \ldots, \widehat{g}_n) = \{g \in \mathfrak{M}(X) / g(\{x_i\}) = \widehat{g}_i, \ i = 1, 2, \ldots, n\}$$

is the class of monotone measures of $\mathfrak{M}(X)$ with coinciding values on singletons.

Analogously, let

$$\mathfrak{M}_c(X;\widehat{g}_1,\widehat{g}_2,\dots,\widehat{g}_n)=\mathfrak{M}_c(X)\cap\mathfrak{M}(X;\widehat{g}_1,\widehat{g}_2,\dots,\widehat{g}_n)$$

be the class of Choquet capacities of order two with the same property and let

$$\mathfrak{M}_\lambda(X;\widehat{g}_1,\widehat{g}_2,\dots,\widehat{g}_n)=\mathfrak{M}_\lambda(X)\cap\mathfrak{M}(X;\widehat{g}_1,\widehat{g}_2,\dots,\widehat{g}_n)$$

be the same class for λ-additive measures. It is clear that

$$\mathfrak{M}_\lambda(X;\widehat{g}_1,\widehat{g}_2,\dots,\widehat{g}_n)\subset\mathfrak{M}_c(X;\widehat{g}_1,\widehat{g}_2,\dots,\widehat{g}_n),$$

where $\lambda>-1$ is a free parameter of the distribution of the λ-additive monotone measure $g_\lambda\in\mathfrak{M}_\lambda(X;\widehat{g}_1,\widehat{g}_2,\dots,\widehat{g}_n)$ with the normalization condition (2.75). If $\sum_{i=1}^{n}\widehat{g}_i=1$, then it can be assumed that $\lambda_0\equiv 0$ (which means that g_0 is a probability measure); otherwise λ is a root of the polynomial

$$\begin{aligned}\Pi(\lambda)=&\left(\prod_{i=1}^{n}\widehat{g}_i\right)\lambda^{n-1}+\dots+\left(\sum_{i<j<k}\widehat{g}_i\widehat{g}_j\widehat{g}_k\right)\lambda^2\\&+\left(\sum_{i<j}\widehat{g}_i\widehat{g}_j\right)\lambda+\sum_{i=1}^{n}\widehat{g}_i-1.\end{aligned}\tag{2.85}$$

Let $L\equiv\{\lambda_1,\lambda_2,\dots,\lambda_l\}$ be the set of real roots of (2.85) $(\lambda>-1)$. Let $L\neq\varnothing$. We introduce the following notation:

$$\mathfrak{M}^L(X)=\{g_{\lambda_i}\in\mathfrak{M}_\lambda(X;\widehat{g}_1,\widehat{g}_2,\dots,\widehat{g}_n)\,|\,\lambda_i\in L,\ \ i=1,2,\dots,l\}.$$

It is clear that the values $\widehat{g}_i^*$ are not "free" (for any $\lambda\in L$):

$$\widehat{g}_i^*=1-\frac{1}{\lambda}\left\{\prod_{j=1\,j\neq i}(1+\lambda\widehat{g}_j)-1\right\},\quad i=1,2,\dots,n,$$

and if $\lambda>0$, then $\widehat{g}_i\le\widehat{g}_i^*$, $i=1,\dots,n$; if $-1<\lambda<0$, then $\widehat{g}_i\ge\widehat{g}_i^*$, $i=1,\dots,n$.

Analogously, we can construct the classes $L^*=\{\lambda^*>1\,|\,\lambda^*=-\frac{\lambda}{1+\lambda},\ \lambda\in L\}$, $\mathfrak{M}^{L^*}(X)\subset\mathfrak{M}^*(X;\widehat{g}_1^*,\widehat{g}_2^*,\dots,\widehat{g}_n^*\}$.

The classes

$$\mathfrak{P}^L(X)=\{P_{g_\lambda}\in\mathfrak{P}(X)\,|\,\lambda\in L\},\quad \mathfrak{P}^{L^*}(X)=\{P_{g^*_{\lambda^*}}\in\mathfrak{P}(X)\,|\,\lambda^*\in L^*\}$$

are the classes of probability measures. We calculate the distances as follows:

$$D_2(\mathfrak{P}^L(X),\mathfrak{M}^L(X)) = \min_{\lambda'\lambda''\in L} D_2(P_{g_{\lambda'}},g_{\lambda''}) = \min_{\lambda\in L} D_2(P_{g_\lambda},g_\lambda)$$

$$= \min_{\lambda\in L}\sqrt{\frac{1}{n!}\sum_{\sigma\in S_n}\sum_{i=1}^{n}\widehat{g}_i^2\left\{\prod_{j=1}^{i(\sigma)-1}\{1+\lambda\widehat{g}_{\sigma(j)}\} - \frac{1}{n!}\sum_{\tau\in S_n}\prod_{k=1}^{i(\sigma)-1}\{1+\lambda\widehat{g}_{\tau(k)}\}\right\}^2}, \tag{2.86}$$

$$D_2(\mathfrak{P}^{L^*}(X),\mathfrak{M}^{L^*}(X)) = \min_{\lambda^*\in L^*} D_2(g^*_{\lambda^*},P_{g^*_{\lambda^*}})$$

$$= \min_{\lambda^*\in L^*}\sqrt{\frac{1}{n!}\sum_{\sigma\in S_n}\sum_{i=1}^{n}\widehat{g}_i^{*2}\left\{\prod_{j=1}^{i(\sigma)-1}\{1+\lambda^*\widehat{g}^*_{\sigma(j)}\} - \frac{1}{n!}\sum_{\tau\in S_n}\prod_{k=1}^{i(\sigma)-1}\{1+\lambda^*\widehat{g}^*_{\tau(k)}\}\right\}^2}. \tag{2.86'}$$

Let these distances be reached on the monotone measures $g_{\widehat{\lambda}}$ and $g^*_{\widehat{\lambda}'}$:

$$D_2(\mathfrak{P}^L(X),\mathfrak{M}^L(X)) = D_2\left(g_{\widehat{\lambda}},P_{g_{\widehat{\lambda}}}\right),$$

$$D_2(\mathfrak{P}^{L^*}(X),\mathfrak{M}^{L^*}(X)) = D_2\left(g^*_{\widehat{\lambda}'},P_{g^*_{\widehat{\lambda}'}}\right).$$

Definition 2.15. The pair of monotone measures $g_{\widehat{\lambda}},g^*_{\widehat{\lambda}'}$ is called the λ-additive fuzzy approximation to insufficient expert data $\widehat{g}_i$, $i=1,\dots,n$.

Definition 2.16. The pair of probability measures $P_{g_{\widehat{\lambda}}},P_{g^*_{\widehat{\lambda}'}}$ is called the probability approximation to insufficient expert data $\widehat{g}_i$, $i=1,\dots,n$.

Note that if $\sum_{i=1}^{n}\widehat{g}_i=1$ and we know that g_0 is a probability measure, then the problem of measure restoration does not exist at all. If we know that g_0 is not a probability measure, then it is assumed that $\lambda\neq 0$.

It is not difficult to verify that $(g_\lambda)^* = g_{-\lambda/(1+\lambda)} \equiv g^*_{\lambda^*}$, $\lambda^* = -\frac{\lambda}{1+\lambda}$; when $\lambda\to 0$ in L, we have $\lambda^*\to 0$ in L^*, and the minima in (2.86), (2.86′) are respectively obtained on $\widehat{\lambda}$, $|\widehat{\lambda}| = \min_{\lambda\in L}|\lambda|$, and λ^* because $\min_{\lambda^*\in L^*}|\lambda^*| = \min_{\lambda\in L}\frac{|\lambda|}{1+\lambda} = \frac{|\widehat{\lambda}|}{1+\widehat{\lambda}} = |\widehat{\lambda}^*|$.

We obtain $(\widehat{\lambda})^* = \widehat{\lambda}'$, $(g_{\widehat{\lambda}})^* = g^*_{\widehat{\lambda}^*} = g^*_{\widehat{\lambda}'}$. This result can be presented in the form of the following proposition.

Proposition 2.21. *The probability approximation corresponds to the approximation of dual monotone measures and* $D_2(\mathfrak{M}^L,\mathfrak{P}^L) = D_2(\mathfrak{M}^{L^*},\mathfrak{P}^{L^*})$.

Proposition 2.22. *The probability measures of the probability approximation coincide.*

Proof. Assume that for the dual monotone measures, $g_{\widehat{\lambda}} \le g^*_{\widehat{\lambda}^*}$. It is clear that $g_{\lambda_1}\le g_{\lambda_2}$ if $\lambda_1\le\lambda_2$ $(\lambda_1,\lambda_2\in L)$ and

$$g_{\widehat{\lambda}}\le P_{g_{\widehat{\lambda}}},\quad P_{g^*_{\widehat{\lambda}^*}}\le g^*_{\widehat{\lambda}^*}.$$

The monotone measures $g_{\widehat{\lambda}}, g^*_{\widehat{\lambda}^*}$ are the nearest ones to probability measures in terms of the distance D_2. Thus we see that it is only the probability distribution that corresponds to the case $\lambda = 0$, $P_{g_{\widehat{\lambda}}} = P_{g^*_{\widehat{\lambda}^*}}$. □

As is known [9], the distance between two monotone measures coincides with that between their dual measures. Since $(\widehat{g}_\lambda)^* = g^*_{\widehat{\lambda}}$ and $(P_{g_{\widehat{\lambda}}})^* \equiv P_{g^*_{\widehat{\lambda}}}$, we have

$$D_2(\mathfrak{M}^L, \mathfrak{P}^L) = D_2(\mathfrak{M}^{L^*}, \mathfrak{P}^{L^*}).$$

As may happen in reality, insufficient data of dual monotone measures (in the form of Table 2.2) can be provided not by one but several experts $\mathbb{E}_X = \{I_1, I_2, \ldots, I_E\}$.

Definition 2.17. We say that data $\widehat{g}_i^\alpha$, $i = 1, \ldots, n$, $\alpha \in \mathbb{E}_X$ (defined by (2.84)), are insufficient expert data of the monotone measure g given by experts $\mathbb{E}_X$.

Insufficient expert data produce classes $\{\mathfrak{M}^{L\alpha}, \mathfrak{M}^{L^*_\alpha}, \mathfrak{P}^{L\alpha}, \mathfrak{P}_{L^*_\alpha}\}$, $\alpha \in \mathbb{E}_X$, from which we can construct the class of probability approximations $\{\widehat{P}_\alpha | \alpha \in \mathbb{E}_X\}$ and the class of λ-additive approximations $\{(\overline{g}_\alpha, \overline{g}^*_\alpha) | \alpha \in \mathbb{E}_X\}$.

Definition 2.18. A pair of dual monotone measures defined for any $A \subseteq X$ as

$$\widetilde{g}(A) = \min_{\alpha \in \mathbb{E}_X} \widehat{P}_\alpha(A), \quad \widetilde{g}^*(A) = \max_{\alpha \in \mathbb{E}_X} \widehat{P}_\alpha(A) \tag{2.87}$$

is called an optimal approximation of zeroth approach.

Definition 2.19. A pair of monotone measures defined for any $A \subseteq X$ as

$$\overline{g}(A) = \min_{\alpha \in \mathbb{E}_X} \overline{g}_\alpha(A), \quad \overline{g}'(A) = \max_{\alpha \in \mathbb{E}_X} \overline{g}^*_\alpha(A)$$

is called an optimal approximation of first approach.

Definition 2.20. A pair of monotone measures defined for any $A \subseteq X$ as

$$\overline{\overline{g}}(A) = \max_{\alpha \in \mathbb{E}_X} \overline{g}_\alpha(A), \quad \overline{\overline{g}}''(A) = \min_{\alpha \in \mathbb{E}_X} \overline{g}^*_\alpha(A)$$

is called an optimal approximation of second approach.

Definition 2.21. Distances $D_2(g, \widetilde{g}) = D_2(g^*, \widetilde{g}^*)$, $D_2(g, \overline{g}) = D_2(g^*, \overline{g}^*)$, and $D_2(g, \overline{\overline{g}}) = D_2(g^*, \overline{\overline{g}}^*)$ are respectively called optimal approximation errors of zeroth, first, and second approaches.

Proposition 2.23. *The monotone measures of the first (second) optimal approximation are dual monotone measures.*

Proof. For any $A \subseteq X$, $\overline{g}'(A) = \max\limits_{\alpha\in\mathbb{E}} \overline{g}^*_\alpha(A) = \max\limits_{\alpha\in\mathbb{E}}(1-\overline{g}_\alpha(\overline{A})) = 1 - \min\limits_{\alpha\in\mathbb{E}} \overline{g}_\alpha(\overline{A} = 1 - \overline{g}(\overline{A})$, i.e., $\overline{g}' = \overline{g}^*$. Analogously, we obtain $\overline{\overline{g}}' = \overline{\overline{g}}^*$. □

Proposition 2.24. *The following inequalities hold between optimal approximations of zeroth, first, and second approaches:*

1. $$\overline{g}(A) \le \overline{\overline{g}}(A); \quad \overline{\overline{g}}^* \le \overline{g}^*(A),$$
2. $$\overline{g}(A) \le \widetilde{g}(A); \quad \widetilde{g}^* \le \overline{g}^*(A),$$
3. $$\overline{\overline{g}}(A) \le \widetilde{g}^*(A); \quad \widetilde{g}^* \le \overline{\overline{g}}^*(A).$$

Proof. For any $A \subseteq X$, we have

1. $$\overline{g}(A) = \min_{\alpha\in\mathbb{E}_X} \overline{g}_\alpha(A) \le \max_{\alpha\in\mathbb{E}_X} \overline{g}_\alpha(A) = \overline{\overline{g}}(A);$$
$$\overline{\overline{g}}^*(A) = \min_{\alpha\in\mathbb{E}_X} \overline{g}^*_\alpha(A) \le \max_{\alpha\in\mathbb{E}_X} \overline{g}^*_\alpha(A) = \overline{g}^*(A),$$
2. $$\overline{g}(A) = \min_{\alpha\in\mathbb{E}_X} \overline{g}_\alpha(A) \le \min_{\alpha\in\mathbb{E}_X} \mathbb{P}_{\overline{g}_\alpha}(A) = \widetilde{g}(A);$$
$$\widetilde{g}^*(A) = \max_{\alpha\in\mathbb{E}_X} \mathbb{P}_{\overline{g}^*_\alpha}(A) = \max_{\alpha\in\mathbb{E}_X} \mathbb{P}_{\overline{g}_\alpha}(A) \le \max_{\alpha\in\mathbb{E}_X} \overline{g}^*(A) = \widetilde{g}^*(A),$$
3. $$\overline{\overline{g}}(A) = \max_{\alpha\in\mathbb{E}_X} \overline{g}_\alpha(A) \le \max_{\alpha\in\mathbb{E}_X} \mathbb{P}_{\overline{g}^*_\alpha}(A) = \widetilde{g}^*_\alpha(A);$$
$$\widetilde{g}(A) = \min_{\alpha\in\mathbb{E}_X} \mathbb{P}_{\overline{g}_\alpha}(A) = \min_{\alpha\in\mathbb{E}_X} \mathbb{P}_{\overline{g}^*_\alpha}(A) \le \min_{\alpha\in\mathbb{E}_X} \overline{g}^*(A) = \overline{\overline{g}}^*(A).$$

The proposition is proved. □

Comparing (2.87) with the representation of Choquet capacities of order two through the associated probabilities (Sect. 2.3.3), we clearly see that between them there exists a certain relationship. The insufficiency of data on distribution is a source of fuzziness, which fact is reflected in the class of expert associated probability measures $\{\widehat{\mathbb{P}}_\alpha | \alpha \in \mathbb{E}_X\}$. According to (2.76) and (2.77), to represent any monotone measure by probabilities, it is necessary and sufficient to have exactly $n!$ probability distributions and their relations to permutations on the set X. Then it becomes obvious that the monotone measure induced by probability distributions $\{\widehat{\mathbb{P}}_\alpha | \alpha \in \mathbb{E}_X\}$ must be related to Choquet capacities of order two.

As defined in Sect. 2.3.3, the dual monotone measures $\widetilde{g}'$, $\widetilde{g}^*$ are called an optimal approximation of zeroth approach. Their connection with Choquet capacities of order two can be described as follows. If $\mathrm{Card}(\mathbb{E}_X) = 1$ (the case with one expert), then $\widetilde{g} = \widetilde{g}^*$ and an optimal approximation is a probability. There are $n!$ associated probabilities that formally coincide, which can be represented as a class filled up with one and the same probability $\widetilde{g}$.

If $1 < \mathrm{Card}(\mathbb{E}_X) \le n!$, then $n!$ associated probabilities $\widetilde{g}$ (as well as $\widetilde{g}^*$) can be formally represented as probabilities of the form $\{\widehat{\mathbb{P}}_\alpha | \alpha \in \mathbb{E}_X\}$, while other probabilities $n! - \mathrm{Card}(\mathbb{E}_X)$ are chosen arbitrarily again from the same set $\{\widehat{\mathbb{P}}_\alpha | \alpha \in \mathbb{E}_X\}$. Clearly, the set of $n!$ probabilities thus filled up is the set of associated probabilities of the Choquet capacity of order two defined by formula (2.87): for any $A \subseteq X$,

$$\widetilde{g}(A) = \min_{\sigma \in S_n} \mathbb{P}_\sigma(A) = \min_{\alpha \in \mathbb{E}_X} \widehat{\mathbb{P}}_\alpha(A),$$

$$\widetilde{g}^*(A) = \max_{\sigma \in S_n} \mathbb{P}_\sigma(A) = \max_{\alpha \in \mathbb{E}_X} \widehat{\mathbb{P}}_\alpha(A),$$

where $\widetilde{g}$ and $\widetilde{g}^*$ are respectively the lower and the upper capacities of order two.

Note that to determine the lower and the upper Choquet capacities of order two, their associated probabilities are to be filled up arbitrarily with elements of the class $\{\widehat{\mathbb{P}}_\alpha | \alpha \in \mathbb{E}_X\}$. Such an arbitrary filling up gives the unique pair of dual monotone measures—lower and upper Choquet capacities of order two.

If $\mathrm{Card}(\mathbb{E}_X) > n!$, then every $n!$ probabilities chosen from the class $\{\widehat{\mathbb{P}}_\alpha | \alpha \in \mathbb{E}_X\}$ define Choquet capacities of order two. Altogether, we obtain $K = C_N^{n!}$ $(N = \mathrm{Card}(\mathbb{E}_X))$ dual pairs of capacities. Denote them as follows:

$$\widetilde{g}_1, \widetilde{g}_2, \ldots, \widetilde{g}_K \quad (\widetilde{g}_1^*, \widetilde{g}_2^*, \ldots, \widetilde{g}_K^*).$$

Proposition 2.25. *An optimal approximation of zeroth approach $(\widetilde{g}, \widetilde{g}^*)$ is written for any $A \subseteq X$ as*

$$\widetilde{g}(A) = \min_{1 \le i \le K} \widetilde{g}_i(A), \quad \widetilde{g}^*(A) = \max_{1 \le i ` K} \widetilde{g}_i^*(A). \tag{2.88}$$

Proof. We will prove the first equality of (2.88) (the second one is proved quite similarly).

For any $A \subseteq X$, let

$$\widetilde{g}(A) = \min_{\alpha \in \mathbb{E}_X} \widehat{\mathbb{P}}_\alpha(A).$$

Note that for the capacities $\widetilde{g}_i$, $i = 1, 2, \ldots, k$, whose associated probabilities $\mathbb{P}_\sigma^{(i)}$, $\sigma \in S_n$, are

$$\{P_\sigma^{(i)}\}_{\sigma \in S_n} \subset \{\widehat{\mathbb{P}}_\alpha | \alpha \in \mathbb{E}_X\},$$

we have

$$\widetilde{g}_i(A) = \min_{\sigma \in S_n} \mathbb{P}_\sigma^{(i)}(A) \ge \min_{\alpha \in \mathbb{E}_X} \widehat{\mathbb{P}}_\alpha(A) = \widetilde{g}(A).$$

Then

$$\min_{1 \le i \le K} \widetilde{g}_i(A) \ge \widetilde{g}(A). \tag{2.89}$$

Clearly, $\exists \alpha_0 \in \mathbb{E}_X$, so that $\widetilde{g}(A) = \widehat{\mathbb{P}}_{\alpha_0}(A)$ and there exists at least one Choquet capacity of order two $\widetilde{g}_{i_0}$ whose associated probabilities belong to $\widehat{\mathbb{P}}_{\alpha_0}$. Then

$$\widetilde{g}_i(A) = \widehat{\mathbb{P}}_{\alpha_0}(A) \ge \min_{\alpha \in \mathbb{E}_X} \widehat{\mathbb{P}}_\alpha(A) = \min_{\sigma \in S_n} \mathbb{P}_\sigma^{(i_0)}(A) = \widetilde{g}_{i_0}(A),$$

i.e.,

$$\min_{1\leq i\leq K}\widetilde{g}_i(A)\leq\widetilde{g}(A). \tag{2.90}$$

Comparing (2.89) and (2.90), we obtain (2.88). Hence we can come to the following conclusion: the dual measures $\widetilde{g}$, $\widetilde{g}^*$ of optimal approximation are Choquet capacities of order two for $\mathrm{Card}(\mathbb{E}_X)\leq n!$; they are respectively minimal and maximal elements of the definite class of Choquet capacities of order two for $\mathrm{Card}(\mathbb{E}_X)>n!$. □

The constructed three variants of optimal approximations obtained from insufficient data on the partly unknown monotone measures g, g^* represent their "restored images."

2.3.4 Conclusions

The insufficiency of expert-objective data on the distribution of a population is a source of possibilistic uncertainty. In this case, one cannot use the classical numerical characteristics of the central tendency (a sampling average, median, modes, and others) for statistical analysis, which, naturally, must be replaced by fuzzy statistics such as FEV, ME, and so on (see Sect. 2.4). To evaluate them, one must construct, on the one hand, the corresponding monotone measure, and on the other hand, the fuzzy subset. Both these objects have the ability to contain condensed fuzzy information in fuzzy average values that are the most typical characteristics of the population.

The fuzzy dual measures $(\widetilde{g},\widetilde{g}^*)$ restored from insufficient expert data are respectively the minimal and the maximal elements of the definite class of Choquet capacities of order two.

2.4 Insufficient Expert Data and Fuzzy Averages

Three new versions of the most typical value (MTV) [38, 106] of a population (generalized weighted averages) are introduced. The first version, $WFEV_g$, is a generalization of the weighted fuzzy expected value ($WFEV$) [40] for any monotone measure g on a finite set, and it coincides with the $WFEV$ when a sampling probability distribution is used. The second and the third versions are respectively the weighted fuzzy expected intervals $WFEI$ and $WFEI_g$, which are generalizations of the $WFEV$, namely, MTVs of the population for a sampling distribution and for any monotone measure g on a finite set, respectively, when the fuzzy expected interval (FEI) [107] exists but the FEV [107] does not. The construction process is based on the Friedman–Schneider–Kandel (FSK) [40] principle and results in new MTVs called the $WFEI$ and the $WFEI_g$ when the combinatorial interval extension of a function [22, 91] is used.

In the present section, generalizations of $WFEV_g$ and $WFEI_g$—correspondingly $GWFEV_g$ and $GWFEI_g$—are introduced for any monotone measure space. Furthermore, the generalized weighted fuzzy expected value (GWFEV) was expressed in terms of two monotone expectation (ME) [8] values with respect to Lebesgue measure on $[0,1]$. The convergence of iteration processes is caused by a free choice of "weight" function, which is very useful in practical situations. In interval extension ($GWFEI_g$) the combinatorial interval extension of function [22] was successfully used, which is clearly seen in examples. Several examples of new weighted values are discussed, where in many cases they give better estimations than classic estimators with central tendencies such as mean, median, or fuzzy "classic" estimators FEV, FEI, ME.

2.4.1 Introduction

There are two classical approaches to the study of inexact data. When experimental data are sufficiently exact, then probabilistic–statistical methods are used to process them and estimate their general characteristics. If data presented are sufficiently inexact and have intervals, then the methods of the theory of errors can be successfully applied.

However, there are cases in which neither the probabilistic–statistical methods nor those of the theory of errors provide satisfactory results. Then one has to investigate the nature of the means (description, measurement, scaling, etc.) by which data were received.

If data are represented in intervals, their distribution is obscure, they overlap, and are described or obtained by an individual expert (insufficient expert data), then they are considered to be of a combined nature. In that case, along with probabilistic–statistical uncertainty, there arises the so-called possibilistic uncertainty produced by an individual (expert) and demanding the application of fuzzy analysis methods. In such situations, only probabilistic–possibilistic analysis can provide satisfactory results through the use of the fuzzy methods to be discussed below.

In describing such data functionally, in many real situations the property of additivity remains unrevealed for a measurable representation of a set, and this creates an additional restriction. Hence, to study subjective insufficient expert data it is frequently better to use monotone estimators instead of additive ones [5, 19, 24, 56, 57, 85, 95, 96, 100, 103, 118–124, 141, 161, 162, and others].

Let us, for example, consider three typical symptoms x_1, x_2, x_3, which indicate some illness y. Let an expert (physician) provide objective–subjective data using his/her wide experience and medical records of patients (another expert would certainly provide different data).

Assume that we have the following information: 80% of patients with illness y exhibit the symptoms x_1 and x_2, while 20% of them have the symptoms x_1 and x_3. This information can be written using the monotone instead of the additive measure g defined on the subsets of the set $X = \{x_1, x_2, x_3\}$ (Table 2.3).

Table 2.3 Distribution table showing dual monotone measures g and g^*

$A \subseteq X$	$g(A)$	$g^*(A) = 1 - g(\overline{A})$
$\{x_1\}$	0	1
$\{x_2\}$	0	0.8
$\{x_3\}$	0	0.2
$\{x_1, x_2\}$	0.8	1
$\{x_1, x_3\}$	0.2	1
$\{x_2, x_3\}$	0	1
$\{x_1, x_2, x_3\}$	1	1

We have already mentioned that monotone measures were first used in fuzzy analysis in the 1980s by Sugeno [141]. Since any measure is connected with integral calculus, along with measurable functions, he also constructed the integral of a measurable function. It is called Sugeno's integral for a compatibility function of a fuzzy subset with respect to a monotone measure and is also known as the FEV. Later, Kandel [56] called it fuzzy statistics.

The fuzzy integral is a functional that assigns some number or a compatibility value to each fuzzy subset when the monotone measure is already fixed. As is known [141], the concept of a fuzzy integral makes it possible to condense information provided by a compatibility function and a monotone measure. Having the monotone measure determined, we can estimate a fuzzy subset by the most typical compatibility value (MTV). The MTV is essentially different in content and significance from a probabilistic average even when a probabilistic measure is used instead of a monotone measure. The preimage of the MTV with respect to a compatibility function distinguishes the most typical representative elements of the considered fuzzy subset from the rest of the universe.

In this section, we discuss the main estimators of fuzzy statistics: the FEV of the population, the fuzzy expected interval (FEI), and the WFEV [40, 56, 57, 107, 134, 135].

As is known, fuzzy averages differ both in form and content from probabilistic–statistical averages and other numerical characteristics such as *mode* and *median*. Nevertheless, in some cases, "nonfuzzy" (objective) and "fuzzy" (subjective) averages coincide [57]. For a given set of fuzzy subsets with compatibility function values from the interval $[0,1]$, the fuzzy average determines the most typical representative compatibility value (FEV) or the interval of compatibility values (FEI).

Fuzzy statistics plays an essential role in probabilistic–possibilistic analysis and is most effectively used in fuzzy expert (decision-making) systems. In the case of fuzzy data, fuzzy averages (FEV) are mainly constructed on population groups, but when these data are insufficient, instead of FEV we use the FEI.

It is important to note that the FEV seldom satisfies the requirements for the most typical value (MTV). For a sampling distribution of the population, Friedman et al. [40] constructed a process for calculating the WFEV. This process is based on a two-postulate principle (FSK). According to these authors, the $WFEV$ is the most typical value of a compatibility function ($MTV = WFEV$).

In this section, the $WFEV_g$ is a calculational process employing the probabilistic representation of a monotone measure on a finite set, i.e., the so-called class of associated probabilistic distributions [8], which during the representation of inexact data enables one to estimate associated probabilities by intervals of belief and thus to determine monotone measure values.

Therefore, a monotone measure can be represented (estimated) by intervals, and for this, it is necessary to use the interval extension of functions in $WFEI$ or $WFEI_g$. In that case, we no longer face the problem of uncertainty of fuzzy distribution. We think that the use of the $WFEI$ needs further research, which, in turn, will open new perspectives for fuzzy data processing when data are insufficient and their distribution is obscure.

We have developed software for estimating weighted values presented in this section.

In the second part of this section (Sects. 2.4.7–2.4.8) we formulate the definitions of the GWFEV and GWFEI and investigate their properties. Examples illustrating the applications of these statistics are also discussed.

Different authors believe the MTV to be different fuzzy values. We will not discuss this topic here but only mention that these values are FEV, FEI, $WFEV$, ME, $WFEVg$, $GWFEV$, etc. In our opinion, there cannot be any preferable MTV in fuzzy statistics, since all of them are expert estimators and give better or worse results depending on a specific problem. Our objective here is to carry out analysis in order to establish which FEV gives a better representation under given circumstances.

2.4.2 Monotone Measure and the FEV

Definition 2.22 ([106]). Let $(X,\mathscr{B})$ be a measurable space, $\mathscr{B}$ a Borel field (σ-algebra). A function $g : \mathscr{B} \Longrightarrow [0,1]$ is called a monotone measure if the following conditions are fulfilled:

(i) $g(\oslash)=0$, $g(X)=1$.
(ii) If $A \subset B$ and $A,B \in \mathscr{B}$, then $g(A) \leq g(B)$.
(iii) If $\{A_k/1 \leq k < \infty\}$ is a monotone sequence $\forall A_k \in \mathscr{B}$, then $\lim_{k\to\infty} g(A_k) = g\left(\lim_{k\to\infty} A_k\right)$.

$(X,\mathscr{B},g)$ is called a monotone measure space.

Let $A(\cdot)$ be a compatibility function of the fuzzy subset A, and let $A(\cdot): X \to [0,1]$ be a $\mathscr{B}$-measurable function, i.e., $\forall\alpha \in [0,1]$, $H_\alpha = \{x \in X/A(x) \geq \alpha\} \in \mathscr{B}$.

Definition 2.23 ([57]). The FEV of the compatibility function $A(\cdot)$ of a fuzzy subset A with respect to the monotone measure g is Sugeno's integral over X:

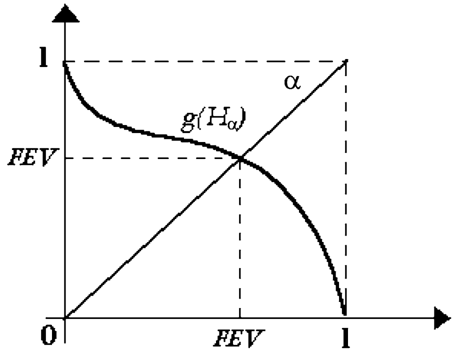

Fig. 2.4 Geometric concept of calculating the FEV

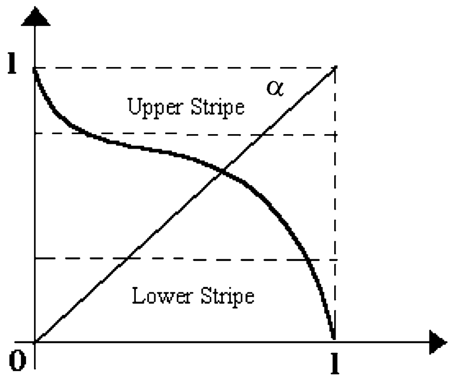

Fig. 2.5 FEV cuts "upper" and "lower" strips of $g(H_\alpha)$

$$FEV(A(\cdot)) \equiv \oint_X A(x) \circ g(\cdot) \overset{\Delta}{=} \sup_{\alpha \in [0,1]} \{\alpha \wedge g(H_\alpha)\}, \tag{2.91}$$

where $\wedge$ denotes the minimum of two arguments.

If $g(H_\alpha)$, $\alpha \in [0,1]$, is a continuous function, then the geometric interpretation of the FEV is as shown in Fig. 2.4.

It clearly follows that the FEV somehow "averages" the values of the compatibility function $A(\cdot)$ not in the sense of a statistical average but by cutting subsets of the α level, whose values of a monotone measure g are either sufficiently "high" or sufficiently "low."

Thus the FEV gives a concrete value of the compatibility function $A(\cdot)$, this value being the most typical characteristic of all possible values with respect to the monotone measure g, obtained by cutting off the "upper" and "lower" strips on the graph of $g(H_\alpha)$ (Fig. 2.5).

Thus the information carried by $A(\cdot)$ and g gets condensed in the FEV, which is the most typical value of all compatibility values.

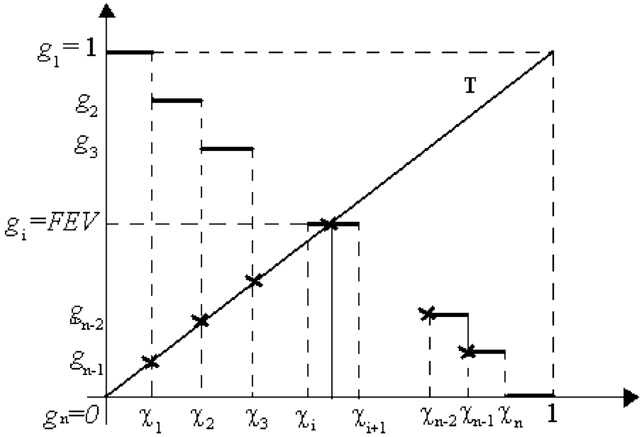

Fig. 2.6 Geometric concept of calculating the FEV

Consider the situation in which $X = \{x_1, x_2, \ldots, x_n\}$ is a finite set ordered so that $A(x_1) \le A(x_2) \le \cdots \le A(x_n)$. Set $X_i = \{x_i, \ldots, x_n\}$, $i = 1, 2, \ldots, n$. As is known [57], the FEV can be calculated by the formula

$$FEV(A(\cdot)) = \max_i \{A(x_i) \wedge g(X_i)\} = \min_i \{A(x_i) \vee g(X_i)\}, \qquad (2.92)$$

where $\vee$ is the maximum of two arguments. If $\chi_i \equiv A(x_i)$, $g_i \equiv g(X_i)$, then a possible geometric interpretation of (2.92) is as shown in Fig. 2.6.

Below we will consider some interesting examples of the calculation of the FEV.

Example 2.1 ([56]). The following statistical data were collected in Beersheba, Israel. In the course of 55 years after 1920, the maximum temperatures recorded there on July 1 were the following:

51 days	90–92°F (average 91°F)
1 day	106°F,
1 day	122°F,
1 day	124°F,
1 day	132°F.

The problem is to determine the value of the high temperature (“hot weather”) in this city on July 1 and to answer the question as to what high temperature is characteristic of Beersheba on this particular day.

It is understood that the base variable “hot weather” is a fuzzy subset of temperature distribution on the whole population. For one expert living in the south, the weather is hot when the temperature is higher than 80°F, while for another expert living in the north, a hot-weather temperature is somewhere below 80°F.

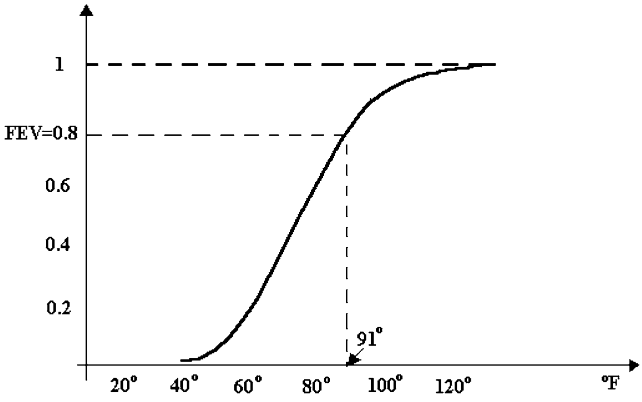

Fig. 2.7 Compatibility curve for "hot weather"

This is the reason why the notion of "hot weather" is fuzzy and given by the function constructed by some expert. Suppose the compatibility curve is as shown in Fig. 2.7).

This problem is first solved by classical statistics using the probabilistic $mean = (91 \cdot 51 + 484)/55 = 93.2$°F, $median = 91$°F. Clearly, the mean cannot describe the typical representative temperature of "hot weather" on July 1 because naturally, it must coincide with the median (because with a high-frequency (51 out of 55) "high temperature" varies from 90 to 92°F). In that case, fuzzy statistics proposes to use the FEV. If the sampling distribution is used instead of the monotone measure g (there is no other information available about g), then $g(H_\alpha) = card(H_\alpha)/55$, where $card$ is the cardinality of the set H_α. The FEV is calculated using (2.92): $FEV(A(\cdot)) = 0.8$, which means that the temperature is $A^{-1}(0.8) = 91$°F, i.e., according to the expert who evaluates "hot weather" by the compatibility curve shown in Fig. 2.7, the most typical representative temperature of hot weather on July 1 is 91°F.

If the expert changes and his/her compatibility function is more "southern" (Fig. 2.8), then $FEV(A(\cdot)) = 0.01543, A^{-1}(0.01543) = 110$°F, while $mean(A(\cdot)) = 0.0235$, i.e., $mean = 94$°F, which in fact is very "low" according to the southern expert.

It can be said that the FEV is a subjective expert characteristic for the population. According to expert data, the FEV is the most typical representative value among the compatibility values of the fuzzy subset.

Example 2.2. Let the base variable be "high salary," creating some fuzzy subset on the set of employees. Consider the salary earned by a number of people and the subjective (expert) compatibility values for χ shown in the following table:

$$1 \text{ person earns } 3.00 \to \chi = 0.40$$
$$3 \text{ person earns } 4.00 \to \chi = 0.50$$

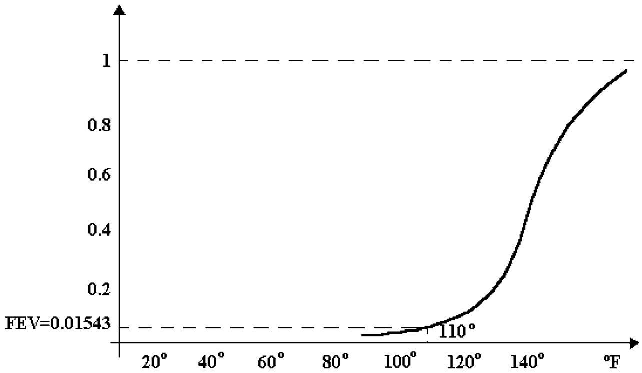

Fig. 2.8 Compatibility curve for "hot weather" ("southern")

$$4 \text{ person earns } 4.20 \to \chi = 0.55$$
$$2 \text{ person earns } 4.50 \to \chi = 0.60$$
$$2 \text{ person earns } 10.00 \to \chi = 1.00$$

Suppose that the following statistical data are available to calculate the FEV:

Group #	x_i	n_i	χ_i	$n^{(i)}$	$g_i = n^{(i)}/n$	$\chi_i \wedge g_i$
1	3.00	1	0.4	12	1	0.4
2	4.00	3	0.5	11	11/12	0.5
3	4.20	4	0.55	8	8/12	0.55
4	4.50	2	0.6	4	4/12	0.33
5	10.00	2	1.0	2	2/12	0.16

where n_i is the number of people in the ith group, $n^{(i)} \equiv \sum_{j=1}^{n} n_j$, $i = 1,2,\ldots,n$, $n = 5$.

As in the previous example, the sampling distribution is taken instead of the monotone measure g on the whole population. Then $FEV = 0.55$, which coincides with the median $A(0.55) = 4.2$ (A. Kandel showed that in such situations the FEV coincides with the median); i.e., the typical high salary for the whole population is 4.2.

If data are received in extreme situations, then the FEV does not provide a "logical" expected value, because in that case, the information available on the population is assumed to be insufficient. Let us consider the next example.

Example 2.3. Let the compatibility function for the variable "old" be

$$\chi(x) = \begin{cases} 0, & x < 0, \\ x/100, & 0 \le x \le 100, \\ 1, & x > 100, \end{cases}$$

and let the statistical distribution of population groups be as follows:

10 people are [10–20] years old,
25 people are 30 years old,
15 people are 40 years old,
35 people are [45–55] years old,
20 people are [60–70] years old.

As in Example 2.2, the table of statistical data is as shown below:

Group #	x_i	n_i	χ_i	$n^{(i)}$	$g_i = n^{(i)}/n$
1	$[10;20]$	10	$[0.1;0.2]$	100	1.00
2	30	25	0.3	90	0.90
3	40	15	0.4	55	0.65
4	$[45;55]$	35	$[0.45;0.55]$	50	0.50
5	$[60;70]$	20	$[0.6;0.7]$	20	0.20

It is obvious that the FEV cannot be calculated by these data, and if we proceed as in Example 2.1 (where for the interval [90–92°F] the average 91°F is taken), then the result will be unsatisfactory, because information will be lost, and this will lead to a significant decrease in the information entropy.

By introducing interval algebra, Schneider and Kandel [107] proposed a new way in which the operations $\wedge$ (minimum) and $\vee$ (maximum) are defined on intervals, and the procedure of calculating the FEV (on a finite set) is generalized. This method is called the fuzzy expected interval (FEI).

2.4.3 Fuzzy Expected Interval

The concept of the FEI as a method was developed to overcome inaccurate fuzzy information in calculating the FEV. Naturally, the FEI must give the same results as the FEV when intervals are one-point sets. To construct the FEI we introduce the operations max and min of interval algebra. Let us recall the definitions and results from [107] (without their proofs):

Definition 2.24. If $S = [\underline{s}, \overline{s}]$ and $R = [\underline{r}, \overline{r}]$ are intervals, then

$$\max\{S,R\} = S \text{if} \forall s \in S : \exists \widetilde{r} \in R \text{such that} s > \widetilde{r},$$

$$\min\{S,R\} = S \text{if} \forall s \in S : \exists \widetilde{r} \in R \text{such that} s < \widetilde{r}. \tag{2.93}$$

Proposition 2.26. *If $S \cap R = \emptyset$, then*

$$\max\{S,R\} = \begin{cases} R \text{ if } \underline{r} > \overline{s}, \\ S \text{ if } \underline{s} > \overline{r}, \end{cases}$$
$$\min\{S,R\} = \begin{cases} R \text{ if } \overline{r} < \underline{s}, \\ S \text{ if } \overline{s} < \underline{r}. \end{cases} \tag{2.94}$$

Proposition 2.27. *If $S \cap R = \emptyset$, $S \subseteq R$, and $R \not\subseteq S$, then*

$$\max\{S,R\} = \begin{cases} R \text{ if } \overline{r} > \overline{s}, \\ S \text{ if } \overline{s} > \overline{r}, \end{cases}$$
$$\min\{S,R\} = \begin{cases} R \text{ if } \overline{s} > \overline{r}, \\ S \text{ if } \overline{r} > \overline{s}. \end{cases} \tag{2.95}$$

Definition 2.25. Suppose $S \subseteq R$. Then $\exists T\,(T = [\underline{t};\overline{t}])$ suvh that
$\max\{S,R,T\} = T$ if $\forall t \in T : \exists\widetilde{s} \in S$ such that $t \geq \widetilde{s}$ and $\exists\widetilde{r} \in R$ such that $t \geq \widetilde{r}$;
$\min\{S,R,T\} = T$ if $\forall t \in T : \exists\widetilde{s} \in S$ such that $t \geq \widetilde{s}$ and $\exists\widetilde{r} \in R$ such that $t \leq \widetilde{r}$.

Proposition 2.28. *If $R \subseteq S$, then*

$$\max\{R,S\} = [\underline{r};\overline{s}], \quad \min\{R,S\} = [\underline{s};\overline{r}]. \tag{2.96}$$

Definition 2.26. Suppose R and S are any intervals from $\Im(\Re)$ (a set of all intervals on the real numbers $\Re$). One can say that S is "larger" than R if $\overline{s} \geq \overline{r}$.

Thus we can define the operations $\wedge$ and $\vee$ on any interval. Now Example 2.3 can be written in the form

$$FEI = \max\{[0.1;0.2], 0.3, 0.4, [0.45;0.5], 0.2\} = [0.45;0.5],$$

where $[0.1;0.2] = \min\{[0.1;0.2],1\}$, $0.3 = \min\{0.3,0.9\}$, $0.4 = \min\{0.4,\ 0.65\}$, $[0.45;0.5] = \min\{[0.45;0.55],0.5\}$, $0.2 = \min\{[0.6;0.7],0.2\}$, but $\chi^{-1}([0.45;0.5]) = [45;50]$. Therefore the most typical age of the considered population with respect to the variable "old" is the interval $[45;50]$.

One may come across examples in which the information available for the frequency distribution of the population is scarce and inexact and the group frequencies are represented by intervals.

Example 2.4. Consider the base variable "old" with the same compatibility function as in Example 1.1.3. The population consists of two groups:

Group #	x_i	n_i	χ_i	$n^{(i)}$	g_i
1	15	[10; 15]	0.15	?	?
2	20	[20; 30]	0.20	?	?

This means that in the first group, 10–15 children are 15 years old, while in the second group, 20–30 children are 20 years old. What is the MTV in this case?

Generally speaking, the values of the monotone measure g_i are intervals whose upper and lower bounds are calculated by [107]

$$\underline{g}_j = \frac{\sum_{i=1}^{k} \min\{\underline{n}_i;\overline{n}_i\}}{\sum_{i=j}^{k} \min\{\underline{n}_i;\overline{n}_i\} - \sum_{i=1}^{j-1} \max\{\underline{n}_i;\overline{n}_i\}},$$

$$\overline{g}_j = \frac{\sum_{i=1}^{k} \max\{\underline{n}_i;\overline{n}_i\}}{\sum_{i=j}^{k} \max\{\underline{n}_i;\overline{n}_i\} - \sum_{i=1}^{j-1} \min\{\underline{n}_i;\overline{n}_i\}}, \tag{2.97}$$

where k is the number of groups in the whole population and $[\underline{n}_i;\overline{n}_i] \equiv n_i$ are frequency intervals of the ith group. If formulas (2.97) are used, then the intervals are given by $g_i = \left[\underline{g}_i;\overline{g}_i\right]$, where $i = 1,2$ are calculated so that $\underline{g}_1 = \overline{g}_1 = 1$, $\underline{g}_2 = 20/(10+30) = 0.25$, $\overline{g}_2 = 30/(10+30) = 0.75$, and the table looks like this:

Group #	x_i	n_i	χ_i	$n^{(i)}$	g_i
1	15	[10; 15]	0.15	[30; 45]	[1; 1]
2	20	[20; 30]	0.20	[20; 30	[0.25; 0.75].

Then $FEI = \max\{\min(0.15,1),\min(0.2,[0.25;0.75])\} = \max\{0.15,0.2\} = 0.2$ and $\chi^{-1}(0.2) = 20$, which means that the most typical group in the whole population is the second one.

In many cases, information is more uncertain than in the above examples and is represented in terms of linguistic variables [182] such as "almost," "more or less," "more," and "much more." In every problem, the subject (expert) constructs a table of relationships for each of the linguistic variables, where the linguistic variables are converted into frequency intervals (mapping table):

Linguistic variable	Lower border	Upper border
Almost	$x-10\%$	$x-1\%$
More or less	$x-10\%$	$x+10\%$
Much more	$2x$	$+\infty$

Note that while receiving data, each linguistic variable creates some population group with a frequency interval. In this case, the FEI has already been calculated.

An example of how to calculate the FEI by means of one expert system of decision-making is given below. In this example, the general system of decision-making is as follows:

"If the condition is fulfilled, then act."

Let us consider a situation for the population when a decision must be made on raising someone's salary:

If "the earning is high," then "raise the salary."

More specifically:

If an earning is ≥ 5, then the salary must be raised by 1%.

Suppose the information on population groups is as follows:

30 people, more or less, earn \$2.5,
50 people earn \$[4–5],
70–100 people earn \$5.5,
50–70 people earn \$[7–8].

The following question arises: Does this population of employees get an increase in salary?

Let the compatibility function of the base variable "high salary" be

$$\chi(x) = \begin{cases} 0, & x < 0, \\ x/10, & 0 \leq x \leq 10, \\ 1, & x > 10. \end{cases}$$

The first population group is created by the linguistic variable "more or less." The above mapping table gives the frequency interval

$$[30 - 10\% \text{ of } 30; 30 + 10\% \text{ of } 30] = [27, 33],$$

and thus we obtain the following distribution table:

Group #	$[\underline{x}_i;\overline{x}_i]$	$n_i = [\underline{n}_i;\overline{n}_i]$	$\chi_i = \left[\underline{\chi}_i;\overline{\chi}_i\right]$	$g_i = \left[\underline{g}_i;\overline{g}_i\right]$
1	2.5	27–33	0.25	1
2	4.0–5.0	50	0.4–0.5	0.84–0.89
3	5.5	70–100	0.55	0.55–0.68
4	7.0–8.0	50–70	0.7–0.8	0.24–0.28

Then $FEI = \max\{\min(0.25, 1), \min([0.4;0.5], [0.84;0.89]), \min(0.55, [0.55; 0.68]), \min([6.7;0.8], [0.24;0.28])\} = \max\{0.25, [0.4;0.5], 0.55, [0.24;0.9]\}$ or $FEI = 0.55$, but $\chi^{-1}(FEI) = \chi^{-1}(0.55) = 5.5 = \chi^{-1}(MTV)$. Since $\chi^{-1}(MTV) > 5.05$, we can say that the employees get their salaries raised.

Despite the fact that the FEV gives a good representation of the most typical population group ($MTPG$) (when data are sufficient) and the FEI gives an interval estimation of the MTV of the compatibility curve (when data on the population groups are insufficient), yet in some cases both the FEV and the FEI give unsatisfactory results. Such a case is illustrated by the next example.

Example 2.5. Suppose one has obtained the following table of compatibility values:

Group # of group	n_i	χ_i	g_i	$\max(\chi_i, g_i)$
1	70	0.05	1	0.05
2	30	0.3	0.3	0.3

If one chooses $FEV = 0.3$ as the most representative value of the function χ, then the group of 70% frequency with the compatibility value 0.05 is ignored. The mean of 0.125 is also unsatisfactory. It would be better to take into consideration the following two factors when calculating the MTV [40]:

1. The MTV must consider groups with a higher frequency in the whole population.
2. The MTV must consider how close it is to the groups with high compatibility values.

Note that these factors are conditional and vary depending on a subjective opinion about the MTV. But it should be said a priori that both factors have played an essential role in the development of the new method to be presented in the next section.

2.4.4 Weighted Fuzzy Expected Value

Friedman et al. proposed a new scheme for calculating the MTV [40], which is based on a two-factor principle. Consider, for example, the following two population groups:

Group #	χ	n
i	χ_i	n_i
j	χ_j	n_j

Suppose $n_i > n_j$. Then:

1. Population effectiveness: the MTV must be "less far" from χ_i than from χ_j, since $n_i > n_j$.
2. The effective location of the MTV with respect to compatibility values: the distance between the MTV and the compatibility value of the ith group $|\chi_i - MTV|$ participates in the definition of the MTV with weight values proportional to $w(|\chi_i - MTV|)$, where w is a strictly decreasing function.

Suppose a variational sampling $\sim\begin{pmatrix}(x_1, x_2, \ldots, x_k)\\(n_1, n_2, \ldots, n_k)\end{pmatrix}$ is given, $\chi_i = A(x_i)$ are the compatibility values of some fuzzy set $A \subset X = \{x_1, x_2, \ldots, x_k\}$, $w(x)$ is a nonnegative monotonically decreasing function defined over the interval $[0,1]$, and $l > 1$ is a real number. Consider the following equation with respect to s:

$$s = \frac{\chi_1 w(|\chi_1 - s|)n_1^l + \chi_2 w(|\chi_2 - s|)n_2^l + \cdots + \chi_k w(|\chi_k - s|)n_k^l}{w(|\chi_1 - s|)n_1^l + w(|\chi_2 - s|)n_2^l + \cdots + w(|\chi_k - s|)n_k^l}. \tag{2.98}$$

Definition 2.27. The solution of (2.98) is called the WFEV of order l with the attached weight function w of compatibility values $(\chi_1,\ldots,\chi_k)$ $(MTV \equiv WFEV(A(\cdot),w))$.

The parameter l measures the dependence of frequencies of population groups on the $WFEV$. The rate at which the function w decreases defines the "closeness" of the $WFEV$ to higher compatibility values of χ_i. By virtue of the above-mentioned two-factor principle, the mapping of weighting (2.98) is invariant with respect to the MTV, which is the fixed point of the mapping. The authors of [38] use the function $w(x) = e^{-\lambda x}$ $(\lambda > 0)$ instead of w. Specifically, for a pair (l,λ), the values are $l = 2$, $\lambda = 1$. To solve (2.98), they use the iteration method $s_n = f(s_{n-1})$, where $s_0 = FEV$ (the function f is the value on the right-hand side of (2.98)), and after three or four iterations they obtain an accuracy of $\varepsilon = 10^{-3}$. Let us return to some of the above examples of the use of the $WFEV$. In the case of Example 1.1.5, if $l = 2$ and $\lambda = 1$, then $WFEV \approx 0.083$, $FEV = 0.3$, $mean = 0.125$, $median = 0.05$. Clearly, the FEV and the median ignore the groups with 70% and 30% frequency, respectively. The mean is close to the compatibility value with a higher frequency but represents the measure of "typicality" of the population more insufficiently than the $WFEV$. The latter uses the two-factor principle and is the most typical value for the population. According to the authors of [40], $MTV = WFEV$.

Example 2.6. The population consists of four groups with the following table of compatibility values:

# of group	χ_i	n_i	g_i
1	0.125	7	1
2	0.375	19	0.93
3	0.625	31	0.74
4	0.875	43	0.43

If $l = 2$ and $\lambda = 1$, then $FEV = 0.625$, $mean = 0.65$, $median = 0.625$, $mode = 0.875$, $WFEV = 0.745$. As in the previous example, the mean is a "better" MTV than $FEV = median$, but "worse" than $mode = 0.875$. This is best summarized in the $WFEV$, and so $MTV = WFEV$.

Example 2.7. The population consists of three groups with the following table of compatibility values:

Group #	χ_i	n_i	g_i
1	0.2	35	1
2	0.3	25	0.65
3	0.6	40	0.4

Then $FEV = 0.4$, $mean = 0.385$, $median = 0.3$, $WFEV = 0.402$, $(l = 2, \lambda = 1)$, $mode = 0.6$.

Clearly, neither the *mean* nor the *median* is a sufficient MTV. The *mean* is slightly better than the *median*, the FEV is better than the *mean*, and the $WFEV$ is much better than both because it is closer to the compatibility value of a higher-frequency group and also takes into account the existence of groups 1 and 2 with 60% frequency.

2.4.5 Weighted Fuzzy Expected Interval

It is important to note that it is impossible to calculate the FEV when data on the population groups are insufficient. Hence a method for calculating the FEI was developed that effectively uses the operations $\vee$ and $\wedge$ from interval algebra. This iteration process is stable, and for one-point intervals the FEI coincides with the FEV. Naturally, a problem like the one discussed in Sect. 2.4 arises during the calculational process of the $WFEV$ when the starting point of the iteration process $s_n = f(s_{n-1})$ cannot be found, but the FEI exists. How can the FEI be used to construct a similar process? Here we construct a new iteration process using interval analysis, where the essential base components are the FEI and principles of $WFEV$ construction.

Suppose the variational sampling $\sim \begin{pmatrix} (x_1, x_2, \ldots, x_k) \\ (n_1, n_2, \ldots, n_k) \end{pmatrix}$ is given, $\chi_i = \chi_{\widetilde{A}}(x_i)$ are the compatibility values of some fuzzy set $\widetilde{A} \subset X = \{x_1, x_2, \ldots, x_k\}$, and n_i and χ_i are intervals: $n_i = [\underline{n}_i; \overline{n}_i]$, $\chi_i = \left[\underline{\chi}_i; \overline{\chi}_i\right]$, $i = 1, 2, \ldots, k$. Let $w(x)$ be a nonnegative monotonically decreasing function defined over the interval $[0,1]$, and $l > 1$ a real number.

Definition 2.28. The weighted fuzzy expected interval ($WFEI$) of order l with the attached weight function w of compatibility values $\{\chi_1, \ldots, \chi_k\}$ is called the limit of the iteration process of the combinatorial interval extension [22]:

$$s_n = \frac{\sum_{i=1}^{k} w\left(\left|\left[\underline{\chi}_i; \overline{\chi}_i\right] - s_{n-1}\right|\right) \cdot \left[\underline{n}_i^l; \overline{n}_i^l\right] \cdot \left[\underline{\chi}_i; \overline{\chi}_i\right]}{\sum_{i=1}^{k} w\left(\left|\left[\underline{\chi}_i; \overline{\chi}_i\right] - s_{n-1}\right|\right) \cdot \left[\underline{n}_i^l; \overline{n}_i^l\right]}, \tag{2.99}$$

where $s_0 \equiv FEI$.

It is denoted by $WFEI(A(\cdot), w)$. It is clear that $WFEI$ is an interval extension of $WFEV$ when FEV does not exist, but FEI does.

Below we formulate an essential proposition that unifies all the weighted means presented in this paper and retains the correctness of generalization of the statistical notions.

Proposition 2.29. *If $FEV = FEI$ and the intervals of compatibility values χ_i and frequencies n_i are one-point intervals, then*

$$WFEV(A(\cdot), w) = WFEI(A(\cdot), w).$$

Clearly, the proof is not difficult. Note that the property of compression of the function w is sufficient for the iteration process (2.99) to converge.

2.4.6 Weighted Fuzzy Expected Value with Respect to the Monotone Measure (WFEV$_{gl}$)

Suppose the variational sampling $\sim\begin{pmatrix}(x_1, x_2, \ldots, x_k)\\(n_1, n_2, \ldots, n_k)\end{pmatrix}$ is given, $\chi_i = A(x_i)$ are the compatibility values of some fuzzy set A. Let w be a nonnegative monotonically decreasing function defined over the interval $[0,1]$, and let $l > 1$ be a real number.

Equation (2.98) can be rewritten as

$$s = \frac{\chi_1 w(|x_1 - s|)\left(\frac{n_1}{k}\right)^l + \chi_2 w(|x_2 - s|)\left(\frac{n_2}{k}\right)^l + \cdots + \chi_k w(|x_k - s|)\left(\frac{n_k}{k}\right)^l}{w(|x_1 - s|)\left(\frac{n_1}{k}\right)^l + w(|x_2 - s|)\left(\frac{n_2}{k}\right)^l + \cdots + w(|x_k - s|)\left(\frac{n_k}{k}\right)^l}. \quad (2.100)$$

Definition 2.29. A monotone measure that for every subset B of the sampling $X = \{x_1, \ldots, x_k\}$ is equal to the lth power of the frequency of B,

$$g^l_{\text{sampling}}(B) \equiv \left(\frac{\sum\limits_{x_i \in B} n_i}{k}\right)^l = \left(\frac{n_B}{k}\right)^l,$$

is called a monotone measure induced by a sampling distribution.

Then

$$g^l_{\text{sampling}}(\{x_i\}) = \left(\frac{n_i}{k}\right)^l, \quad i = 1, 2, \ldots, k. \quad (2.101)$$

It is obvious that during the weighting we consider the values of the measure g^l_{sampling} in (2.100) only on sets of one element (“weights” of sets of one element).

Let $X = \{x_1, \ldots, x_k\}$ be a finite set, $\left(X, 2^X, g\right)$ a monotone measure space, $A(\cdot)$ a compatibility function of the fuzzy subset A, $A(\cdot) : X \to [0;1]$ $(\chi_i = A(x_i))$; let w be some "weight" function and $l > 1$ a real number.

By virtue of (2.100) and Definition 2.29, the following two new postulates of constructing the MTV with respect to the monotone measure g on the set X can be formulated, which in what follows are referred to as the Friedman–Schneider–Kandel (FSK) principles:

1. Monotone measure distribution effectiveness: the MTV is "less far" from χ_i than from χ_j if $g(\{x_i\}) > g\left(\{x_j\}\right)$.
2. The effective location of the MTV with respect to compatibility values: the distance between the MTV and the compatibility values χ_i of the element $x_i \in X$): $|\chi_i - MTV|$ participates in the definition of the MTV with weight values proportional to $w(|\chi_i - MTV|)$, where w is a strictly decreasing function.

Similarly to (2.100) and (2.101), let us consider the following equation with respect to s:

$$s = \frac{\sum\limits_{i=1}^{k} \chi_i w\left|(\chi_i - s)\right| g^l(\{x_i\})}{\sum\limits_{i=1}^{k} w\left|(\chi_i - s)\right| g^l(\{x_i\})}. \tag{2.102}$$

Definition 2.30. The solution of (2.102) is called the WFEV of order l with the attached weight function w of the compatibility function χ with respect to the monotone measure g.

It is denoted by $WFEV_{gl}(A(\cdot), w)$, $\left(MTV = WFEV_{gl}\right)$.

On the set $\{1, 2, \ldots, k\}$ there exist $k!$ permutations. Denote any permutation by $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(k))$ and the set of all possible permutations by S_k.

Definition 2.31. If $\sigma \in S_k$ is a permutation, then the probability distribution

$$\begin{gathered}
P_\sigma^{(l)}\left(x_{\sigma(1)}\right) = g^l\left(\left\{x_{\sigma(1)}\right\}\right), \\
P_\sigma^{(l)}\left(x_{\sigma(2)}\right) = g^l\left(\left\{x_{\sigma(1)}, x_{\sigma(2)}\right\}\right) - g^l\left(\left\{x_{\sigma(1)}\right\}\right), \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
P_\sigma^{(l)}\left(x_{\sigma(i)}\right) = g^l\left(\left\{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\right\}\right) - g^l\left(\left\{x_{\sigma(1)}, \ldots, x_{\sigma(i-1)}\right\}\right), \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
P_\sigma^{(l)}\left(x_{\sigma(k)}\right) = 1 - g^l\left(\left\{x_{\sigma(1)}, \ldots, x_{\sigma(k-1)}\right\}\right),
\end{gathered}$$

is called an associated probability distribution of the monotone measure g^l; $\left\{P_\sigma^{(l)}\right\}_{\sigma \in S_k} = \left\{P_\sigma^{(l)}\left(x_{\sigma(1)}\right), \ldots, P_\sigma^{(l)}\left(x_{\sigma(k)}\right)\right\}_{\sigma \in S_k}$ is called the class of associated probabilities of the monotone measure g^l.

It is known [138] that $\forall x_i \subset X$, there exists a permutation $\tau_i \in S_k$ such that

$$g^{\ell}\left(\{x_i\}\right) = P_{\tau_i}^{(l)}\left(x_i\right) \equiv P_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right).$$

Now (2.102) takes the form

$$s = \frac{\sum\limits_{i=1}^{k} \chi_i w\left|(\chi_i - s)\right| P_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right)}{\sum\limits_{i=1}^{k} w\left|(\chi_i - s)\right| P_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right)}. \tag{2.103}$$

This is the probabilistic representation of the $WFEV_{gl}$ by associated probabilities $P_{\tau_1}^{(l)}, P_{\tau_2}^{(l)}, \ldots, P_{\tau_n}^{(l)}$ of the monotone measure g.

Obviously, we can construct the iteration process for (2.103) as we have done for (2.98):

$$s_n = \frac{\sum\limits_{i=1}^{k} \chi_i w\left|(\chi_i - s_{n-1})\right| P_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right)}{\sum\limits_{i=1}^{k} w\left|(\chi_i - s_{n-1})\right| P_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right)},$$

where $s_0 = FEV\left(A(\cdot)\right)$.

Let the values χ_i and $P_{\tau_i}^{(l)}(\cdot)$ be intervals: $\chi_i = \left[\underline{\chi}_i; \overline{\chi}_i\right]$, $P_{\tau_i}^{(l)} = \left[\underline{P}_{\tau_i}^{(l)}; \overline{P}_{\tau_i}^{(l)}\right]$; let w be a nonnegative monotonically decreasing function defined over the interval [0; 1], and let $l > 1$ be a real number.

Definition 2.32. The weighted fuzzy expected interval $WFEI_{gl}$ of order l with the attached weight function w of the compatibility function $A(\cdot)$ with respect to the monotone measure g is called the limit of the iteration process of the combinatorial interval extension

$$s_n = \frac{\sum\limits_{i=1}^{k} \left[\underline{\chi}_i; \overline{\chi}_i\right] w\left(\left|\left[\underline{\chi}_i; \overline{\chi}_i\right] - s_{n-1}\right|\right) \left[\underline{P}_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right); \overline{P}_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right)\right]}{\sum\limits_{i=1}^{k} w\left(\left|\left[\underline{\chi}_i; \overline{\chi}_i\right] - s_{n-1}\right|\right) \left[\underline{P}_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right); \overline{P}_{\tau_i}^{(l)}\left(x_{\tau_i(1)}\right)\right]}, \tag{2.104}$$

where $s_0 = FEV\left(A(\cdot)\right)$. It is denoted by $WFEI_{gl} = WFEV_{gl}\left(A(\cdot), w\right)$.

It is clear that $WFEI_{gl}$ is an interval extension of the $WFEV_{gl}$ and the following propositions are true.

Proposition 2.30. *If $FEV = FEI$ and the intervals of compatibility values χ_i and the values of associated probabilities (or the values of the monotone measure g) $P_\tau(\cdot)$ are one-point intervals, then*

$$WFEI_{gl} = WFEV_{gl}.$$

Clearly, the proof is not difficult.

Proposition 2.31. *If $X = \{x_1, \ldots, x_k\}$ is the set of variational sampling*

$$\sim \begin{pmatrix} (x_1, x_2, \ldots, x_k) \\ (n_1, n_2, \ldots, n_k) \end{pmatrix}$$

and $g : 2^X \to [0;1]$ is the "sampling" monotone measure

$$g = g_{\text{sampling}},$$

then the following generalized WFEVs coincide:

$$WFEV_{gl} = WFEV, WFEI_{gl} = WFEI.$$

Clearly, the proof is not difficult.

Conclusion. When data on the population groups are scarce, the process of fuzzy statistical estimation consists of two stages. First, the generalization of the fuzzy weighted estimator follows from a small amount of information. The fuzzy weighted estimator is formally constructed by interval analysis and creates an information entropy growth. Second, the flexible FSK principle leads to an entropy decrease of information that is condensed in the generalized fuzzy statistics that are the new MTVs of the population, which are respectively called the weighted fuzzy expected intervals ($WFEI$ and $WFEI_{gl}$) and the WFEV $WFEV_{gl}$ with respect to the monotone measure g.

2.4.7 Generalized Weighted Fuzzy Expected Value and the Generalized Weighted Fuzzy Expected Interval

In Sect. 2.4.6, the notion of WFEV ($WFEV_{g\ell}$) of a compatibility function on a finite set X with respect to the monotone measure g was defined. Let $X = \{x_1, \ldots, x_k\}$ be a finite set, $A(\cdot)$ a compatibility function of the fuzzy subset $A \subset X$, w e a nonnegative strictly decreasing function with values in the interval $[0;1]$, and $\ell > 1$ a real number. Then $WFEV_{g\ell}$ is called a solution of the following equation with respect to s:

$$s = \frac{\sum\limits_{i=1}^{n} \chi_i w(|\chi_i - s|) g^{\ell}(\{x_i\})}{\sum\limits_{i=1}^{n} w(|\chi_i - s|) g^{\ell}(\{x_i\})}, \tag{2.105}$$

where $\chi_i \equiv A(x_i)$. Equation (2.105) is constructed according to the Friedman–Schneider–Kandel (FSK) principle when on the finite set we consider not the probabilistic, but the monotone measure g^{ℓ}. As is known, the FSK principle is based on two postulates. The first postulate concerns effectiveness of distribution of the monotone measure g in the "weighting" procedure. In (2.105) this is represented by "fuzzy weights" $g^{\ell}(\{x_i\})$ of singleton sets $\{x_i\}$, $i = 1,2,\dots,k$. The second postulate states that the MTV has a location such that it is close to the values of "high" compatibility. This is represented in the "weighting process" by weights $w(|\chi_i - MTV|)$. Both postulates are combined into a normalized sum $\chi_i w(|\chi_i - s|)g^{\ell}(\{x_i\})$ presented in (2.105). This means that $MTV \equiv WFEV_{g^{\ell}}$ (the solution of (2.105)) is invariant with respect to the "weight" function constructed by the FSK principle, which is certainly understandable and justified.

As the authors of the FSK principle [40] admit, the two postulates of the "weighting" process cannot be perfect. Some other postulates can also be added. We think that in the case of the first postulate, the "weighting" process has one drawback. The values $g^{\ell}(\{x_i\})$ of the monotone measure g^{ℓ} are considered only on singleton sets $\{x_i\}$, $i = 1,2,\dots,k$, while the values of the same measure on the subsets of X with two or more elements are not. The point is that the values of compatibility $\chi_1, \chi_2, \dots, \chi_k$ are described by the consonant body (theorem of decomposition [182]), i.e., by focal subsets of the form $\{x_i \in X \,|\, A(x_i) \geq \alpha\}$, $0 < \alpha < 1$. But if we consider the second postulate of the FSK principle, weights $w(|\chi_i - MTV|)$, $i = 1,2,\dots,n$, then it becomes obvious that the values of the monotone measure g^{ℓ} on the cuts of the α-level $\{x_i \in X \mid A(x_i) \geq \alpha\}$, $0 \leq \alpha < 1$, participate in the "weighting" process. This strengthens the first postulate. If we consider the condition of normalizing of the "weighting" process, then the invariant form of the "weighting" function is defined by the ratio of two monotone expectations (see (2.108)). All this leads us to the definition of a new weighted fuzzy average that can be defined generally for any monotone measure space.

Suppose X is any universe, and $(X, \mathscr{B}, g)$ is any monotone measure space, $A(\cdot)$ is a $\mathscr{B}$-measurable compatibility function of the fuzzy subset A, $A(\cdot) : X \to [0,1]$. The monotone expectation of $A(\cdot)$ with respect to the monotone measure g is written as follows [8, 56]:

$$E_g(A(\cdot)) = \int_0^1 g\big(\{x \in X \,|\, A(x) \geq \alpha\}\big) d\alpha \stackrel{\text{def}}{=} \int g(H_\alpha) d\alpha, \tag{2.106}$$

where H_α is a fuzzy subset of the α level of $A(\cdot)$; $g(H_\alpha)$ is said to be a measure function of $A(\cdot)$; $d\alpha$ is an element of the Lebesgue measure on $[0,1]$.

Definition 2.33. If $w(\cdot)$ is a positive "weight" function, strictly decreasing, with values in the interval $[0,1]$, and $\ell \geq 1$, $\beta > 0$ are real numbers, then the solution of the equation (with respect to s)

$$s = \frac{\int_0^1 g^{\ell}(\{A^{\beta}(x) w(|A^{\beta}(x) - s|) \geq \alpha\}) d\alpha}{\int_0^1 g^{\ell}(\{w(|A^{\beta}(x) - s|) \geq \alpha\}) d\alpha} \tag{2.107}$$

is called the generalized WFEV of power β of the compatibility function $A(\cdot)$ with weight w and with respect to the monotone measure g^{ℓ}.

We denote this value by

$$GWFEV_{g\ell}^{\beta} \equiv GWFEV_{g\ell}(A^{\beta}(\cdot), w).$$

In terms of the monotone expectation, equation (2.107) can be rewritten as

$$s = E_{ge}\big(A^{\beta}(\cdot) w(|A^{\beta}(\cdot) - s|)\big) / E_{ge}\big(w(|A^{\beta}(\cdot) - s|)\big). \tag{2.108}$$

Proposition 2.32. *If $X = \{x_1, x_2, \ldots, x_k\}$ is a finite set and $\beta > 1$, $\ell > 1$ are real numbers, $\chi_i \equiv A(x_i)$ are values of the compatibility function of a fuzzy subset $A \subset X$, then there exist two probabilistic distributions P_1 and P_2 on X with respect to which (2.108) takes the form of a ratio of mathematical expectations,*

$$s = \frac{E_{P_1}\big(A^{\beta}(\cdot) w(|A^{\beta}(\cdot) - s|)\big)}{E_{P_2}\big(w(|A^{\beta}(\cdot) - s|)\big)}.$$

Proof. If S_k denotes the class of all permutations of the set $\{1, 2, \ldots, k\}$ ($\sigma = (\sigma_{(1)}, \sigma_{(2)}, \ldots, \sigma_{(k)}) \in S_k$), then $\exists \sigma_1 = \sigma_1(s) \in S_k$ and permutations $\sigma_2 = \sigma_2(s) \in S_k$ such that

$$\chi_{\sigma_1(i)}^{\beta} w(|\chi_{\sigma_1(i)}^{\beta} - s|) \le \chi_{\sigma_1(j)}^{\beta} w(|\chi_{\sigma_1(j)}^{\beta} - s|),$$
$$w(|\chi_{\sigma_2(i)}^{\beta}(\cdot) - s|) \le w(|\chi_{\sigma_2(j)}^{\beta} - s|),$$

where $i < j$ and $\chi_{\sigma(i)} \equiv A(x_{\sigma(i)})$. The integrated functions of (2.107) can be written as

$$g^{\ell}\big(\{A^{\beta}(\cdot) \cdot w(|A^{\beta}(\cdot) - s|) \ge \alpha\}\big) = \begin{cases} g_{\sigma_1(1)}^{\ell} & \text{if } 0 \le \alpha < \widehat{\chi}_{\sigma_1(1)}, \\ g_{\sigma_1(2)}^{\ell} & \text{if } \widehat{\chi}_{\sigma_1(1)} \le \alpha < \widehat{\chi}_{\sigma_1(2)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ g_{\sigma_1(k)}^{\ell} & \text{if } \widehat{\chi}_{\sigma_1(k-1)} \le \alpha < \widehat{\chi}_{\sigma_1(k)}, \\ 0 & \text{if } \widehat{\chi}_{\sigma_1(k)} \le \alpha < 1, \end{cases}$$

where

$$g_{\sigma_1(i)}^{\ell} \equiv g^{\ell}\big(\{x_{\sigma_1(1)}, x_{\sigma_1(2)}, \ldots, x_{\sigma_1(i)}\}\big),$$
$$\widehat{\chi}_{\sigma_1(i)} \equiv \chi_{\sigma_1(i)}^{\beta} \cdot w(|\chi_{\sigma_1(i)}^{\beta} - s|), \quad i = 1, 2, \ldots, k,$$

$$g^{\ell}\left(\left\{w(|A^{\beta}(\cdot)-s|)\geq\alpha\right\}\right)=\begin{cases}g^{\ell}_{\sigma_2(1)} & \text{if } 0\leq\alpha<\approx\!\!\!\!\chi_{\sigma_2(1)},\\ g^{\ell}_{\sigma_2(2)} & \text{if } \approx\!\!\!\!\chi_{\sigma_2(1)}\leq\alpha<\approx\!\!\!\!\chi_{\sigma_2(2)},\\ \dots\dots\dots\dots & \\ g^{\ell}_{\sigma_2(k)} & \text{if } \approx\!\!\!\!\chi_{\sigma_2(k-1)}\leq\alpha<\approx\!\!\!\!\chi_{\sigma_2(k)},\\ 0 & \text{if } \approx\!\!\!\!\chi_{\sigma_2(k)}\leq\alpha<1,\end{cases}$$

where

$$g^{\ell}_{\sigma_2(i)}\equiv g^{\ell}\left(\left\{x_{\sigma_2(1)},x_{\sigma_2(2)},\dots,x_{\sigma_2(i)}\right\}\right),\quad \overset{\approx}{\chi}_{\sigma_2(i)}\equiv w(|\chi^{\beta}_{\sigma_2(i)}-s|),\ \ i=1,2,\dots,k.$$

Now (2.107) takes the form

$$s=\frac{\sum\limits_{i=1}^{k}\int\limits_{\widehat{\chi}_{\sigma_1(i-1)}}^{\widehat{\chi}_{\sigma_1(i)}}g^{\ell}_{\sigma_1(i)}d\alpha}{\sum\limits_{i=1}^{k}\int\limits_{\overset{\approx}{\chi}_{\sigma_2(i-1)}}^{\overset{\approx}{\chi}_{\sigma_2(i)}}g^{\ell}_{\sigma_2(i)}d\alpha},$$

i.e.,

$$s=\frac{\sum\limits_{i=1}^{k}g^{\ell}_{\sigma_1(i)}(\widehat{\chi}_{\sigma_1(i)}-\widehat{\chi}_{\sigma_1(i-1)})}{\sum\limits_{i=1}^{k}g^{\ell}_{\sigma_2(i)}(\overset{\approx}{\chi}_{\sigma_2(i)}-\overset{\approx}{\chi}_{\sigma_2(i-1)})},$$

and

$$s=\frac{\sum\limits_{i=1}^{k}\chi_{\sigma_1(i)}\{g^{\ell}_{\sigma_1(i)}-g^{\ell}_{\sigma_1(i+1)}\}}{\sum\limits_{i=1}^{k}\overset{\approx}{\chi}_{\sigma_2(i)}\ \{g^{\ell}_{\sigma_2(i)}-g_{\sigma_2(i+1)}\}}.$$

The differences in braces are the values of the associated probability distributions of the monotone measure g^{ℓ} for permutations $\sigma_1\in S_k$ and $\sigma_2\in S_k$ [135], respectively, and

$$g^{\ell}_{\sigma_j(i)}-g^{\ell}_{\sigma_j(i-1)}\overset{\text{def}}{=}P^{(\ell)}_{\sigma_j}(x_{\sigma_j(i)}),\ \ j=1,2;\ \ i=1,2,\dots,k\ \ (g_{\sigma_j(0)}\equiv 0).$$

Finally, (2.105) can be rewritten as

$$s = \frac{\sum_{i=1}^{k} \chi^{\beta}_{\sigma_1(i)} w(|\chi^{\beta}_{\sigma_1(i)} - s|) \cdot P^{(\ell)}_{\sigma_1}(x_{\sigma_1(i)})}{\sum_{i=1}^{k} w(|\chi^{\beta}_{\sigma_2(i)} - s|) \cdot P^{(\ell)}_{\sigma_2}(x_{\sigma_2(i)})}, \tag{2.109}$$

i.e.,

$$s = \frac{E_{P^{(\ell)}_{\sigma_1}}\left(A^{\beta}(\cdot) w(|A^{\beta}(\cdot) - s|)\right)}{E_{P^{(\ell)}_{\sigma_2}}\left(w(|A^{\beta}(\cdot) - s|)\right)},$$

where $E_{P^{(\ell)}_{\sigma_j}}$ denotes the mathematical expectation with respect to the probability distribution $P^{(\ell)}_{\sigma_j}$, $j = 1, 2$. □

In practice, for a numerical solution of (2.108) or (2.109) the following simple iteration process is used:

$$s^{(N+1)} = f(s^{(N)}), \quad N = 0, 1, \dots, \tag{2.110}$$

where $s^{(0)}$ is the FEV

$$s^{(0)} \equiv FEV_{g\ell}(A^{\beta}(\cdot)) = \int_X A^{\beta}(\cdot) \circ g^{\ell}(\cdot),$$

i.e., for (2.109) we have the iteration process

$$s^{(N+1)} = \frac{\sum_{i=1}^{k} \chi^{\beta}_{\sigma_1^{(N)}(i)} w(|\chi^{\beta}_{\sigma_1^{(N)}(i)} - s^{(N)}|) \cdot P^{\ell)}_{\sigma_1^{(N)}}(x_{\sigma_1^{(N)}(i)})}{\sum_{i=1}^{k} w(|\chi^{\beta}_{\sigma_2^{(N)}(i)} - s^{(N)}|) \cdot P^{\ell)}_{\sigma_2^{(N)}}(x_{\sigma_2^{(N)}(i)})}, \tag{2.111}$$

where the permutations $\sigma_1 = \sigma_1^{(N)}$ and $\sigma_2 = \sigma_2^{(N)}$ depend on the iteration step N and thus

$$f(s^{(N)}) \equiv E_{g\ell}\left(A^{\beta}(\cdot) w(|A^{\beta}(\cdot) - s^{(N)}|)\right) / E_{g\ell}\left(w(|A^{\beta}(\cdot) - s^{(N)}|)\right). \tag{2.112}$$

It can be easily shown that if the function w is chosen effectively [40], the function f has compression in the neighborhood of $s^{(0)}$ and the convergence of (2.112) is guaranteed:

$$\lim_{N \to \infty} s^{(N)} = GWFEV^{\beta}_{g\ell}.$$

Following [40], the function $w(t) = e^{-\lambda t}$ is frequently used in the role of w (where λ is a positive parameter).

Suppose $X = \{x_1, x_2, \ldots, x_k\}$ is a finite set. Let $\chi_j \equiv A(x_i)$ be compatibility values of some fuzzy subset $A \subset X$. Suppose χ_i and g^ℓ are intervals. Then $P_{\sigma_j}^{(\ell)}(\cdot) = g_{\sigma_j}^\ell(\cdot) - g_{\sigma_j}^\ell(\cdot)$ are intervals $(j = 1, 2)$. Set $\chi_i = [\underline{\chi}_i; \overline{\chi}_i]$ and $P_{\sigma_j}^{(\ell)} = [\underline{P}_{\sigma_j}^{(\ell)}; \overline{P}_{\sigma_j}^{(\ell)}]$. Suppose w is a nonnegative, strictly decreasing function; $\ell > 1$, $\beta > 0$ are real numbers and there exists an FEI [107].

Definition 2.34. The generalized weighted fuzzy expected interval $(GWFEI_{g\ell}^{\beta})$ of order ℓ of a compatibility function $A^\beta(\cdot)$ with the attached weight function w with respect to the monotone measure g^ℓ is called the limit of the combinatorial interval extension of the iteration process (2.111):

$$s_{N+1} = \frac{\sum_{i=1}^{k} w\left(\left|[\underline{\chi}_{\sigma_1(i)}; \overline{\chi}_{\sigma_1(i)}] - s_N\right|\right) \cdot [\underline{P}_{\sigma_1(i)}^{(\ell)}; \overline{P}_{\sigma_1(i)}^{(\ell)}] \cdot [\underline{\chi}_{\sigma_1(i)}^{\beta}; \overline{\chi}_{\sigma_1(i)}^{\beta}]}{\sum_{i=1}^{k} w\left(\left|[\underline{\chi}_{\sigma_2(i)}; \overline{\chi}_{\sigma_2(i)}] - s_N\right|\right) \cdot [\underline{P}_{\sigma_2(i)}^{(\ell)}; \overline{P}_{\sigma_2(i)}^{(\ell)}]}, \tag{2.113}$$

where $s_0 \equiv FEI(A^\beta(\cdot))$, $P_{\sigma_j(i)}^{(\ell)} = P_{\sigma_j(i)}^{(\ell)}(N)$, $j = 1, 2$; $i = 1, 2, \ldots, k$; $N = 1, 2, \ldots$.

It is denoted by $GWFEI_{g\ell}^{\beta}(A(\cdot), w)$. Clearly, $GWFEI_{g\ell}^{\beta}$ is the interval extension of $GWFEV_{g\ell}^{\beta}$ when FEV does not exist, but FEI does [107]. Note that for the convergence of the iteration process (2.112), the property of compression of the function w is sufficient. An essential proposition that unifies all the weighted means presented in this section and retains the correctness of generalization of the statistical notions is formulated as the following proposition.

Proposition 2.33. *If the values of compatibility χ_i and associated probabilities $\{P_\sigma^{(\ell)}\}_{\sigma \in S_k}$ are one-point intervals and $FEV = FEI$, then*

$$GWFEV_{g\ell}^{\beta}(A(\cdot), w) = GWFEI_{g\ell}^{\beta}(A(\cdot), w).$$

The proof is obvious.

Conclusion. Sect. 2.4.7 is mainly of descriptive character. The existing FEV and the WFEV are generalized for any monotone measure space $(GWFEV_{g\ell}^{\beta})$. Using the combinatorial interval extension, the generalized weighted fuzzy expected interval $(GWFEI_{g\ell})$ is defined for the case that the construction of an FEV is impossible (but the FEI exists). The obtained new weighted expected values actually represent iteration processes whose convergence is guaranteed by an appropriate choice of a weight function. Propositions on the correctness of generalization are formulated.

Generalized weighted averages can be successfully used in expert systems, where FEV and $WFEV$ are used but cannot be estimated because of data scarcity.

In conclusion, several examples are discussed to illustrate the application of the new fuzzy statistics and to show that generalized weighted expected values often provide better results.

Table 2.4 Associated probabilities of g^{ℓ}

$\sigma \in S_3$	$P_{\sigma}^{(\ell)}(x_1)$	$P_{\sigma}^{(\ell)}(x_2)$	$P_{\sigma}^{(\ell)}(x_3)$
(1,2,3)	$(1/3)^{\ell}$	$(2/3)^{\ell} - (1/3)^{\ell})$	$1 - (2/3)^{\ell}$
(1,3,2)	$(1/3)^{\ell}$	$1 - (2/3 - \varepsilon)^{\ell}$	$(2/3 - \varepsilon)^{\ell} - (1/3)^{\ell}$
(2,1,3)	$(2/3)^{\ell} - (1/3 - \varepsilon)^{\ell})$	$(1/3 - \varepsilon)^{\ell}$	$1 - (2/3)^{\ell}$
(2,3,1)	$1 - (2/3 - \varepsilon)^{\ell}$	$(1/3 - \varepsilon)^{\ell}$	$(2/3 - \varepsilon)^{\ell} - (1/3 - \varepsilon)^{\ell})$
(3,1,2)	$(2/3 - \varepsilon)^{\ell} - (1/3)^{\ell})$	$1 - (2/3 - \varepsilon)^{\ell}$	$(1/3)^{\ell}$
(3,2,1)	$1 - (2/3 - \varepsilon)^{\ell}$	$(2/3 - \varepsilon)^{\ell} - (1/3)^{\ell}$	$(1/3)^{\ell}$

Table 2.5 Distributions of g on X

$B \subset X$	$g(B)$
$\emptyset$	0
$\{x_1\}$	$1/3$
$\{x_2\}$	$1/3 - \varepsilon$
$\{x_3\}$	$1/3$
$\{x_1, x_2\}$	$2/3$
$\{x_1, x_3\}$	$2/3 - \varepsilon$
$\{x_2, x_3\}$	$2/3 - \varepsilon$
$\{x_1, x_2, x_3\}$	1

2.4.8 Examples

Let us consider two examples of the use of the $WFEV_g$ that is a solution of (2.108) ($\beta = 1$, $\ell = 2$, $w(|t|) = e^{-|t|}$, $S_0 = FEV$).

Example 2.8. Suppose $X = \{x_1, x_2, x_3\}$ and F is a fuzzy subset

$$F = \{0.4/x_1, 0.6/x_2, 0.8/x_3\},$$

and also the monotone measure g is given on X, whose associated probabilities are shown in Table 2.4 ($0 < \varepsilon < 1/3$ is a parameter).

Then the distribution of g takes the form shown in Table 2.5.

Now according to (2.111), we obtain the following iteration process for calculating $GWFEV_{g^2}$:

$$s_{N+1} = \frac{0.4e^{-|0.4-s_N|}P_{\sigma_1}^{(2)}(x_1) + 0.6e^{-|0.6-s_N|}P_{\sigma_1}^{(2)}(x_2) + 0.8e^{-|0.8-s_N|}P_{\sigma_1}^{(2)}(x_3)}{e^{-|0.4-s_N|}P_{\sigma_1}^{(2)}(x_1) + e^{-|0.6-s_N|}P_{\sigma_1}^{(2)}(x_2) + e^{-|0.8-s_N|}P_{\sigma_1}^{(2)}(x_3)},$$

where $s_0 \equiv FEV(F(\cdot))$.

By (2.108), the probabilities $P_{\sigma_j}^{(2)}(x_i)$, $j = 1, 2$; $i - 1, 2, 3$, depend on S_N, and for any N we choose different probabilities $P_{\sigma_j}^{(2)}$ from the associated probabilities of the monotone measure g^2 (Table 2.4).

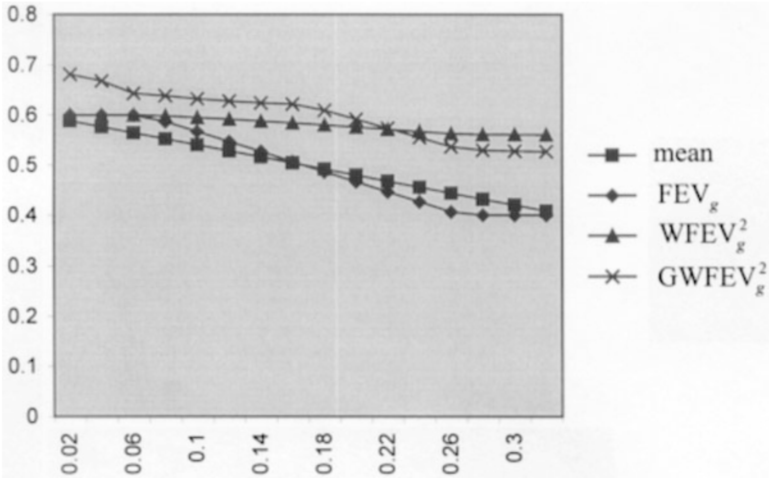

Fig. 2.9 Fuzzy averages and mean

By virtue of (2.105), we obtain the following iteration process for calculating $WFEV_{g^2}$:

$$s_{N+1} = \frac{0.4e^{-|0.4-s_N|}(1/3)^2 + 0.6e^{-|0.6-s_N|}(1/3-\varepsilon)^2 + 0.8e^{-|0.8-s_N|}(1/3)^2}{e^{-|0.4-s_N|}(1/3)^2 + e^{-|0.6-s_N|}(1/3-\varepsilon)^2 + e^{-|0.8-s_N|}(1/3)^2},$$

where $s_0 = FEV$.

Calculations yield

$$FEV(F(\cdot)) = \begin{cases} 0.6 & \text{if } 0 < \varepsilon \le \frac{1}{15}, \\ \frac{2}{3} - \varepsilon & \text{if } \frac{1}{15} < \varepsilon \le \frac{4}{15}, \\ 0.4 & \text{if } \frac{4}{15} < \varepsilon \le \frac{1}{3}. \end{cases}$$

Note that when $\varepsilon = 0$, g is a probabilistic measure (a uniform distribution on X) and $WFEV_g = WFEV$. Because of small values of ε, it makes sense to calculate the *mean*:

$$mean = \sum_i g(\{x_i\})F(x_i) = 0.4 + \left(\frac{1}{3} - \varepsilon\right)0.6.$$

If we give a small pitch to the parameter ε and gather calculations of *mean*, FEV, $WFEV_{g^2}$, $GWFEV_{g^2}$ in one chart (see Fig. 2.9), then it becomes obvious that for $\varepsilon \approx 0$ all three MTVs are close to 0.6 (excepting only $GWFEV_{g^2}$), but $F^{-1}(MTV) = x_2$, and thus x_2 becomes the most typical element of the set X (population). But here a high level of the compatibility of x_3 in $F - 0.8$ is not taken into account. With the growth of ε, the values of *mean* and FEV decrease, which means that $MTV \to 0.4$

Table 2.6 Distributions of g on X

$B \subset X$	$g(B)$
$\emptyset$	0
$\{x_1\}$	1/3
$\{x_2\}$	1/4
$\{x_3\}$	1/2
$\{x_1,x_2\}$	2/3
$\{x_1,x_3\}$	5/6
$\{x_2,x_3\}$	3/4
$\{x_1,x_2,x_3\}$	1

Table 2.7 Associated probabilities of g^ℓ

$\sigma \in S_3$	$P_\sigma^{(\ell)}(x_1)$	$P_\sigma^{(\ell)}(x_2)$	$P_\sigma^{(\ell)}(x_3)$
(1,2,3)	$(1/3)^\ell$	$(2/3)^\ell - (1/3)^\ell)$	$1-(2/3)^\ell$
(1,3,2)	$(1/3)^\ell$	$1-(5/6)^\ell$	$(5/6)^\ell - (1/3)^\ell$
(2,1,3)	$(2/3)^\ell - (1/4)^\ell)$	$(1/4)^\ell$	$1-(2/3)^\ell$
(2,3,1)	$1-(3/4)^\ell$	$(1/4)^\ell$	$(3/4-\varepsilon)^\ell - (1/4)^\ell)$
(3,1,2)	$(5/6)^\ell - (1/2)^\ell)$	$1-(5/6)^\ell$	$(1/2)^\ell$
(3,2,1)	$1-(3/4)^\ell$	$(3/4)^\ell - (1/2)^\ell$	$(1/2)^\ell$

and $F^{-1}(MTV) \to x_1$, or x_1 becomes a new typical element from the population X, which does not consider levels of the compatibility of $x_2 - 0.6$ and $x_3 - 0.8$ in F. We certainly cannot regard this as a good representation. We have a different situation with $WFEV_{g2}$, because $WFEV_{g2} \to 0.56$, i.e., $WFEV_{g2}$ retains closeness to 0.6 though it tends to 0.5, i.e., it takes into account high levels of the compatibility of x_2 and x_3. Obviously, in this case $x_1 \cup x_2$ is the most typical element of the population X, which is better seen in $WFEV_{g2}$ than in the FEV or the *mean*. Note that here the monotone measure g corrects the decision for which the fuzzy averaging has been carried out.

For $\varepsilon \approx 0$, x_3 has the highest compatibility level 0.8, and the distribution of the monotone measure g gives preference almost to none of the elements (Table 2.5). This means that $F^{-1}(MTV)$ must be shifted toward x_3 and the compatibility values of x_1 and x_2 must be considered and therefore $GWFEV_{g2} > WFEV_{g2}$. With the growth of ε, the role of x_3 decreases and the average value becomes $F^{-1}(MTV) = x_2$. When $\varepsilon \to 1/3$, the role of x_2 decreases, the role of x_3 is also slightly weakened, though its compatibility value is the highest one and must be considered anyway. But the role of x_1 increases. This is best of all described in $GWFEV_{g2}$.

Example 2.9. Suppose $X = \{x_1,x_2,x_3\}$ and F is a fuzzy subset

$$F = \{0.4/x_1,\, (0,6-\varepsilon)/x_2,\, 0.8/x_3\}, \quad 0 \le \varepsilon < 0.6,$$

and the distribution of the monotone measure g on X is the associated probabilities of the monotone measure g^ℓ are shown in Tables 2.6 and 2.7.

Calculations indicate that

$$mean = \frac{41}{60} + \frac{1}{4}\varepsilon,$$

$$FEV(F(\cdot)) = \begin{cases} 0.5 & \text{if } 0 \leq \varepsilon \leq 0.1, \\ 0.6-\varepsilon & \text{if } 0.1 \leq \varepsilon \leq 0.2, \\ 0.5 & \text{if } 0.2 \leq \varepsilon \leq 0.6. \end{cases}$$

Let $\beta = 1$; $\ell = 2$. If we give a small pitch to ε, then the iteration process (2.111) takes the form $(s_0 = FEV)$ (numerical estimation of $GWFEV_{g^2}$)

$$s_{N+1} = \frac{0.4e^{-|0.4-s_N|}P^{(2)}_{\sigma_1}(x_1)+(0.6-\varepsilon)e^{-|0.6-\varepsilon-s_N|}P^{(2)}_{\sigma_1}(x_2)+0.8e^{-|0.8-s_N|}P^{(2)}_{\sigma_1}(x_3)}{e^{-|0.4-s_N|}P^{(2)}_{\sigma_1}(x_1)+e^{-|0.6-\varepsilon-s_N|}P^{(2)}_{\sigma_1}(x_2)+e^{-|0.8-s_N|}P^{(2)}_{\sigma_1}(x_3)}.$$

For the numerical estimation of $WFEV_{g^2}$ we have the iteration process

$$s_{N+1} = \frac{0.4e^{-|0.4-s_N|}(1/3)^2+0.6e^{-|0.6-\varepsilon-s_N|}(1/4)^2+0.8e^{-|0.8-s_N|}(1/2)^2}{e^{-|0.4-s_N|}(1/3)^2+e^{-|0.6-\varepsilon-s_N|}(1/4)^2+e^{-|0.8-s_N|}(1/2)^2},$$

where $s_0 = FEV$.

The results are collected in Fig. 2.10.

It is obvious that if $0 \leq \varepsilon \leq 0.1$, then $FEV = 0.5$, and according to FEV, x_2 is the most typical element. Nearly the same result is shown by the *mean* and $WFEV_{g^2}$. But if $0.1 \leq \varepsilon \leq 0.6$, then according to the FEV, x_1 is almost the most typical element with compatibility level 0.5 and fuzzy weight 1/3. Here the FEV does not consider the fuzzy weight of x_2, though its level of compatibility tends to zero. Neither does it consider a high level of compatibility of $x_3 - 0.8$ and high fuzzy weight 0.5. Obviously, the FEV is a bad estimator of the MTV when $0.1 \leq \varepsilon \leq 0.6$. The *mean* is better, though its most typical value is $\{x_1, x_2\}$ when $\varepsilon \to 0.6$. We cannot say the same about the $WFEV_{g^2}$, for which $F^{-1}(MTV) \to x_3$ when $\varepsilon \to 0.6$. It is obvious that the $WFEV_{g^2}$ is better than the FEV and the *mean*.

When $\varepsilon \approx 0$, it is obvious that $F^{-1}(MTV)$ must be shifted toward x_3, because it has the highest compatibility level. If we look at the distribution of g, then $g(\{x_3\}) > g(\{x_i\})$, $i = 1,2$, $g(\{x_3, x_k\}) > g(\{x_1, x_2\})$, $k = 1,2$, i.e., x_3 is represented in the distribution of g with highest weight. Let us consider the $GWFEV_{g^2}$ when $\varepsilon \approx 0$, $F^{-1}(MTV) = x_3$. With the growth of ε, the role of x_3 does not weaken in $F^{-1}(MTV)$, but the role of x_2 does $(F(x_2) \to 0)$. In contrast to the FEV and the *mean*, the $GWFEV_{g^2}$ as well as $WFEV_{g^2}$ do not “forget” the “special role” of x_3 in the MTV, which does not depend on ε, though $F^{-1}(MTV) = \{x_1, x_2\}$.

Example 2.10. Suppose only two groups of the population are given, as shown in Table 2.8.

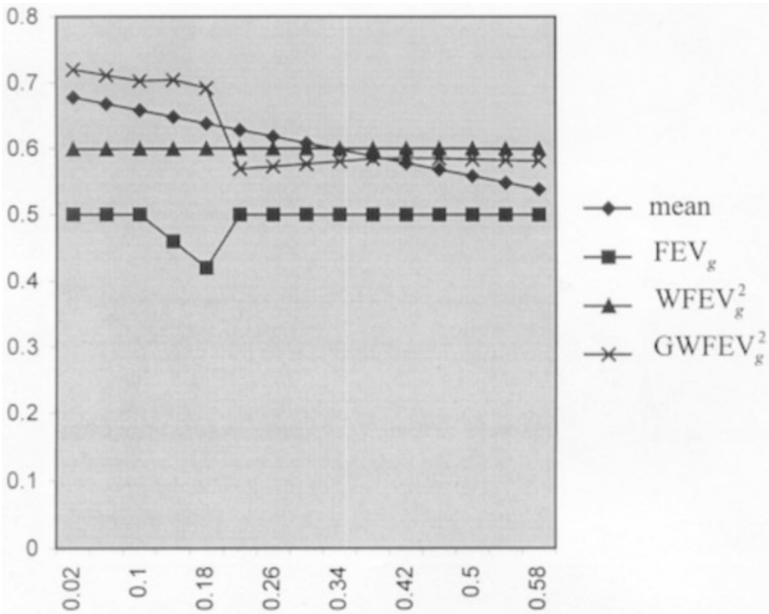

Fig. 2.10 Fuzzy averages and mean

Table 2.8 Associated probabilities of g^ℓ

Group #	x_i	n_i	$F(x_i)$	g_i	σ	$P_\sigma^{(\ell)}(x_1)$	$P_\sigma^{(\ell)}(x_2)$
1	15	[10–15]	$0.15+\varepsilon$	$\left[\frac{1}{4};\frac{5}{7}\right]$	$(1,2)$	$\left[\left(\frac{1}{4}\right)';\left(\frac{5}{7}\right)'\right]$	$\left[1-\left(\frac{5}{7}\right)';1-\left(\frac{1}{4}\right)'\right]$
2	20	[20–30]	$0.25-\varepsilon$	$\left[\frac{4}{7};\frac{3}{4}\right]$	$(2,1)$	$\left[\left(\frac{4}{7}\right)';\left(\frac{3}{4}\right)'\right]$	$\left[1-\left(\frac{3}{4}\right)';1-\left(\frac{4}{7}\right)'\right]$

Here the monotone measure g is a sampling distribution of the population. Since frequencies n_i are intervals, the FEV does not exist. So the FEI is calculated by the procedure described in [107].

We obtain $FEI(F(\cdot)) = 0.2$, $F^{-1}(0.2) = 20$, which means that the second population group is typical. But the FEI does not consider the first group with 15% frequency and compatibility level 0.15 at all. Naturally, the MTV must be slightly shifted toward the first group, which can be done by means of the FSK principle [106]. But in that case, the $WFEV$ cannot be calculated, and we have to use its interval extension (iteration process (2.113)), where $s_0 = FEI$, i.e., the iteration process starts with the interval $s_0 = [0.2;0.2]$. Let $\beta = 1$; $\ell = 2$. By the combinatorial interval extension we get the $(GWFEI_{g^2})$:

$$s_{N+1} = \frac{0.15e^{-|0.15-s_N|}[\underline{P}_\sigma^\ell(x_1);\overline{P}_\sigma^\ell(x_1)] + 0.2e^{-|0.20-s_N|}[\underline{P}_\sigma^\ell(x_2);\overline{P}_\sigma^\ell(x_2)]}{e^{-|0.15-s_N|}[\underline{P}_\sigma^\ell(x_1);\overline{P}_\sigma^\ell(x_1)] + e^{-|0.20-s_N|}[\underline{P}_\sigma^\ell(x_2);\overline{P}_\sigma^\ell(x_2)]}.$$

Table 2.9 Associated probabilities of g^ℓ

Group #	$[\underline{x}_i;\overline{x}_i]$	$[\underline{n}_i;\overline{n}_i]$	$[\underline{\chi}_i;\overline{\chi}_i]$	$g_i = [\underline{g}_i;\overline{g}_i]$
1	2.5	[27–33]	$0.25-\varepsilon$	[0.11–0.16]
2	[4.0–5.0]	50	[0.4+ε–0.5+ε]	[0.20–0.25]
3	5.5	[70–100]	0.55	[0.31–0.44]
4	[7.0–8.0]	[50–70]	[0.7–0.8]	[0.21–0.32]

For the interval estimation of the $WFEI_{g^2}$ we have the iteration process

$$s_{N+1} = \frac{0.15e^{-|0.15-s_N|}[\frac{1}{4};\frac{5}{7}]^2 + 0.2e^{-|0.20-s_N|}[\frac{4}{7};\frac{3}{4}]^2}{e^{-|0.15-s_N|}[\frac{1}{4};\frac{5}{7}]^2 + e^{-|0.20-s_N|}[\frac{4}{7};\frac{3}{4}]^2},$$

where $s_0 = FEI$.

Computer-aided calculations show that the procedure is fast enough and that stability is achieved after three or four iterations:

$$WFEI_{g^2} = [0.169; 0.195].$$

As expected, $WFEI_{g^2} < FEI$, which means that the $WFEI_{g^2}$ considers "possibilities" of the first group. It shows that the most typical element should be the group that, according to its data, is located between the first and second groups. This conclusion is confirmed by $GWFEI_{g^2} = [0.075; 0.474]$, which does not deny the role of x_1 in $F^{-1}(MTV)$ and practically does not make a categorical decision concerning the MTV.

Example 2.11. Let us review the "salary example" considered in Sect. 2.4.3 of this chapter. Here we list only the necessary data. The population consists of four groups. The elements (x_i), their frequencies (n_i) and values of compatibility χ_i are intervals $(0 \le \varepsilon \le 0.15)$, as shown in Table 2.9.

Here $\underline{g}_i = \min \frac{[\underline{n}_i;\overline{n}_i]}{\sum_{j=1}^4 [\underline{n}_j;\overline{n}_j]}$, $\overline{g}_i = \max \frac{[\underline{n}_i;\overline{n}_i]}{\sum_{j=1}^4 [\underline{n}_j;\overline{n}_j]}$, $i = 1,2,3,4$. Whatever the values of ε $(0 \le \varepsilon \le 0.15)$ are, the FEI does not depend on them, and

$$FEI(F(\cdot)) = 0.55,$$

i.e., $F^{-1}(0.55) = 5.5$, and the third population group is the most typical categorically! The FEI does not consider the high characteristics of the second and fourth groups.

The iteration process of the $WFEI_{g^2}$ is

$$\begin{aligned} s_{N+1} = {} & \{e^{-|0.25-\varepsilon-S_N|}(0.25-\varepsilon)g_1^2 + e^{-|[0.4+\varepsilon;0.5+\varepsilon]-S_N|}[0.4+\varepsilon;0.5+\varepsilon]g_2^2 \\ & + e^{-|0.55-S_N|}0.55g_3^2 + e^{-|[0.7;0.8]-S_N|}[0.7;0.8]g_4^2\}/ \\ & /\{e^{-|0.25-\varepsilon-S_N|}g_1^2 + e^{-|[0.4+\varepsilon;0.5+\varepsilon]-S_N|}g_2^2 + e^{-|0.55-S_N|}g_3^2 + e^{-|[0.7;0.8]-S_N|}g_4^2\}, \end{aligned}$$

where $s_0 = FEI = 0.55$.

Table 2.10 Calculations of WFEI and FEI

ε	$WFEI$	FEI
0.01	$[0.470, 0.576]$	0.55
0.02	$[0.465, 0.581]$	0.55
0.03	$[0.468, 0.582]$	0.55
0.04	$[0.471, 0.584]$	0.55
0.05	$[0.474, 0.586]$	0.55
0.06	$[0.476, 0.588]$	0.55
0.07	$[0.479, 0.590]$	0.55
0.08	$[0.483, 0.592]$	0.55
0.09	$[0.485, 0.594]$	0.55
0.1	$[0.487, 0.597]$	0.55
0.11	$[0.488, 0.600]$	0.55
0.12	$[0.490, 0.603]$	0.55
0.13	$[0.492, 0.606]$	0.55
0.14	$[0.494, 0.609]$	0.55
0.15	$[0.496, 0.612]$	0.55

As seen in Table 2.10, the $WFEI_{g^2}$ is sensitive to the growth of ε and is increasing, which means that the share of the second and fourth groups in the definition of the most typical group increases, since the share of the first group decreases. When $\varepsilon \to 0,15$, according to the $WFEI$, the most typical group of the population is $F^{-1}(WFEI_{g^2}) = \{2,3\}$. Although the fourth group also becomes almost the typical one. As we see, the $WFEI_{g^2}$ lacks certainty and chooses several groups as the most typical ones. $WFEI_{g^2}$ is a better estimator for MTV than FEI, which does not consider the role of ε.

Note that the $GWFEI_{g^2}$ is not calculated in this example because the complete distribution of the monotone measure g on the population groups is unknown.

Conclusion. The discussed examples indicate that in many cases it is better to use the $GWFEV_{g^2}$ or the $WFEV_{g^2}$, and if the measure is nonadditive but monotone on the population (monotone measure) and data are represented by intervals, then the only statistic that finds the most typical element of the population is the $WFEI_{g^2}$ or the $GWFEI_{g^2}$.

2.5 Conclusions

As a conclusion we can briefly state that in this chapter, we show that finite probabilistic representations of monotone measure MSR and CBR are in some sense equivalent. Some theorems of monotone measures are proved in MSR analogically to already existing CBR. The probabilistic representations of concrete monotone measures and Dempster–Shafer belief structures are given as separate sections. The main achievement, from an applications point of view, is that CBR has been successfully used in identification problems of monotone measure, based on insufficient expert data. Particular mention should be made of the use of CBR in new constructions of weighted fuzzy averages, whose advantage over other classical fuzzy averages is clearly demonstrated in the large number of examples at the end of the chapter. This represents a new credible mean for defuzzification, which is used in the last, third, part of the book.

Springer

http://www.springer.com/978-1-4614-4250-9

Extremal Fuzzy Dynamic Systems
Theory and Applications
Sirbiladze, G.
2013, XXII, 402 p.,
ISBN: 978-1-4614-4250-9