

Article

# Divergence and Similarity Characteristics for Two Fuzzy Measures Based on Associated Probabilities

Gia Sirbiladze \* , Bidzina Midodashvili and Teimuraz Manjafarashvili

Department of Computer Sciences, Ivane Javakhishvili Tbilisi State University, Tbilisi 0186, Georgia; bidzina.midodashvili@tsu.ge (B.M.); teimuraz.manjafarashvili@tsu.ge (T.M.)

\* Correspondence: gia.sirbiladze@tsu.ge

**Abstract:** The article deals with the definitions of the distance, divergence, and similarity characteristics between two finite fuzzy measures, which are generalizations of the same definitions between two finite probability distributions. As is known, a fuzzy measure can be uniquely represented by the so-called its associated probability class (APC). The idea of generalization is that new definitions of distance, divergence, and similarity between fuzzy measures are reduced to the definitions of distance, divergence, and similarity between the APCs of fuzzy measures. These definitions are based on the concept of distance generator. The proof of the correctness of generalizations is provided. Constructed distance, similarity, and divergence relations can be used in such applied problems as: determining the difference between Dempster-Shafer belief structures; Constructions of collaborative filtering similarity relations; non-additive and interactive parameters of machine learning in phase space metrics definition, object clustering, classification and other tasks. In this work, a new concept is used in the fuzzy measure identification problem for a certain multi-attribute decision-making (MADM) environment. For this, a conditional optimization problem with one objective function representing the distance, divergence or similarity index is formulated. Numerical examples are discussed and a comparative analysis of the obtained results is presented.

**Keywords:** fuzzy measures; associated probabilities; distances; distance generator; divergence indexes; fuzzy measure identification; conditional optimization

**MSC:** 28E10; 03E72; 60A86



**Citation:** Sirbiladze, G.; Midodashvili, B.; Manjafarashvili, T. Divergence and Similarity Characteristics for Two Fuzzy Measures Based on Associated Probabilities. *Axioms* **2024**, *13*, 776. <https://doi.org/10.3390/axioms13110776>

Academic Editors: Hsien-Chung Wu, Ming-Feng Ge and Radko Mesiar

Received: 26 August 2024

Revised: 31 October 2024

Accepted: 5 November 2024

Published: 9 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. On the Information Measures of Distance for Two Probability Distributions

It is known that in statistics or probability theory, as well as in information theory, the concept of distance quantitatively describes the proximity between two statistical objects or the similarity. It can be reduced to two finite probability distributions and/or two statistical samples. Thus, the distance between two populations is interpreted as a measure of closeness between two finite probability distributions. Note that many of the already defined statistical distances are not metrics, and some of them do not have the property of symmetry. Some types of distances that generate so-called squared distances are called (statistical) divergences.

When comparing finite probability distributions, as always Mahalanobis [1], Bhattacharyya [2] and Hellinger [3] distance formulas are used. Of course, depending on the need, the other distances may be used as well. Mahalanobis distance is used in statistical analysis when comparable probability distributions have different means and variances. Bhattacharyya's distance is mainly used in the construction of face recognition processes, and Hellinger's distance is mainly an important tool in the engineering of text mining and classifying documents processes. Note that Hellinger's distance is often called Jeffrey's distance.

Let  $X = \{x_1, \dots, x_n\}$  be a finite set and  $\mathcal{P}(X)$  denote the space of all probability distributions defined on  $X$ . Let  $P^{(1)}$  and  $P^{(2)}$  be any two probability distributions from  $\mathcal{P}(X)$ . Introduce notations

$$P^{(1)}(x_i) = P_i^{(1)} \text{ and } P^{(2)}(x_i) = P_i^{(2)}, \quad i = 1, \dots, n; \quad \sum_{i=1}^n P_i^{(k)} = 1, \quad k = 1, 2.$$

The following expression is called Bhattacharyya’s coefficient of distributions  $P^{(1)}$  and  $P^{(2)}$  [4]:

$$BC(P^{(1)}, P^{(2)}) = \sum_{i=1}^n (P_i^{(1)} \cdot P_i^{(2)})^{1/2}. \tag{1}$$

The Bhattacharyya’s distance between two probability distributions is defined as follows [4]:

$$\mathcal{D}_B(P^{(1)}, P^{(2)}) = -\log(BC(P^{(1)}, P^{(2)})) = -\log\left(\sum_{i=1}^n \sqrt{P_i^{(1)} \cdot P_i^{(2)}}\right). \tag{2}$$

$\mathcal{D}_B(\cdot)$  does not represent a metrics in  $\mathcal{P}(X)$ , but it measures the similarity and closeness between two probability distributions. There is also the Bhattacharyya’s angle [4], which measures the distance between two probability distributions:

$$\Delta_B(P^{(1)}, P^{(2)}) = \arccos(BC(P^{(1)}, P^{(2)})). \tag{3}$$

The Hellinger’s distance between two probability distributions is called an expression [3]:

$$\mathcal{D}_H(P^{(1)}, P^{(2)}) = [1 - BC(P^{(1)}, P^{(2)})]^{1/2} = \sqrt{1 - \sum_{i=1}^n (P_i^{(1)} \cdot P_i^{(2)})^{1/2}}. \tag{4}$$

$\mathcal{D}_H$  forms a bounded metric in  $\mathcal{P}(X)$ , which also measures the similarity and closeness between two probability distributions.

The Mahalanobis distance is a measure of the distance between a point  $\vec{x} = (x_1, \dots, x_n) \in R^n$  and a probability distribution  $P$  on  $R^n$  [1]. Given a probability distribution  $P$  on  $R^n$  with expectation value  $\vec{m} = (m_1, \dots, m_n)$  and positive semi-definite covariance matrix  $C$ . Mahalanobis distance of a point  $\vec{x} = (x_1, \dots, x_n)$  from the probability distribution  $P$  is called an expression

$$\mathcal{D}_M(\vec{x}, P) = \sqrt{(\vec{x} - \vec{m})C^{-1}(\vec{x} - \vec{m})}. \tag{5}$$

1.2. On the Divergences for Two Probability Distributions

Now we turn to the analysis of divergences between two probability distributions. We will give definitions of several divergences based on the concept of distances of probability distributions, which will be important in our further synthesis problems.

In general,  $f$ -divergence is a type of function  $\mathcal{D}_{\nu_f}(P^{(1)}||P^{(2)})$  [5] that measures the difference between two probability distributions  $P^{(1)}$  and  $P^{(2)}$ . However, here we present a slightly different general divergence measure. Divergence was first introduced by A. Rényi [6]:

$$\mathcal{D}_{\nu_R}^{(\alpha)}(P^{(1)} || P^{(2)}) = \frac{1}{\alpha - 1} \log\left(\sum_{i=1}^n \frac{(P_i^{(1)})^\alpha}{(P_i^{(2)})^{\alpha-1}}\right) = \frac{1}{\alpha - 1} \log \mathcal{E}_{P^{(1)}}\left[\left(\frac{P^{(1)}}{P^{(2)}}\right)^{\alpha-1}\right], \tag{6}$$

where  $0 < \alpha < \infty, \alpha \neq 1$ ,  $\mathcal{E}$  is a symbol of mathematical expectation. When we calculate the Rényi divergence for different values of  $\alpha$ , we obtain well-known divergences as particular cases of Rényi divergence.

An important divergence is Kullbak—Leibler divergence [7,8]

$$\mathcal{D}_{V_{KL}}(P^{(1)}||P^{(2)}) = \mathcal{D}_{\nu_R^{(1)}}(P^{(1)} || P^{(2)}) = \sum_{i=1}^n P_i^{(1)} \cdot \log \frac{P_i^{(1)}}{P_i^{(2)}} = \mathcal{E}_{P^{(1)}}(\log \frac{P^{(1)}}{P^{(2)}}), \tag{7}$$

which also represents Shannon’s relative entropy with respect to the distributions  $P^{(1)}$  and  $P^{(2)}$ .  $\mathcal{D}_{V_{KL}}$  is also called an asymmetric divergence.

The following symmetric divergence is produced from it, which today is called the Jeffrey’s divergence [9]:

$$\mathcal{D}_{\nu_J}(P^{(1)}, P^{(2)}) = \mathcal{D}_{\nu_{KL}}(P^{(1)} || P^{(2)}) + \mathcal{D}_{\nu_{V_{KL}}}(P^{(2)} || P^{(1)}) = \mathcal{D}_{\nu_J}(P^{(2)}, P^{(1)}). \tag{8}$$

Then

$$\begin{aligned} \mathcal{D}_{\nu_J}(P^{(1)}, P^{(2)}) &= \sum_{i=1}^n \left( P_i^{(1)} \cdot \log \frac{P_i^{(1)}}{P_i^{(2)}} - P_i^{(2)} \cdot \log \frac{P_i^{(1)}}{P_i^{(2)}} \right) = \\ &= \sum_{i=1}^n (P_i^{(1)} - P_i^{(2)}) \left( \log P_i^{(1)} - \log P_i^{(2)} \right). \end{aligned} \tag{9}$$

The Kullback—Leibler divergence also produces the symmetric Jensen-Shannon divergence [10–12]:

$$\begin{aligned} \mathcal{D}_{\nu_{JS}}(P^{(1)}, P^{(2)}) &= \frac{1}{2}(\mathcal{D}_{\nu_{KL}}(P^{(1)} || R) + \mathcal{D}_{\nu_{KL}}(P^{(2)} || R)) = \\ &= \frac{1}{2} \sum_{i=1}^n \left( P_i^{(1)} \log \left( \frac{2P_i^{(1)}}{P_i^{(1)} + P_i^{(2)}} \right) + P_i^{(2)} \log \left( \frac{2P_i^{(2)}}{P_i^{(1)} + P_i^{(2)}} \right) \right), \end{aligned} \tag{10}$$

where  $R = \frac{1}{2}(P^{(1)} + P^{(2)})$  is a probability distribution on  $\mathcal{P}(X)$ .  $\mathcal{D}_{\nu_{JS}}(\cdot)$  represents a particular case of so-called  $\lambda$ —divergence

$$\mathcal{D}_{\nu_{\lambda}}(P^{(1)}, P^{(2)}) = \lambda \cdot \mathcal{D}_{\nu_{V_{KL}}}(P^{(1)} || R) + (1 - \lambda) \cdot \mathcal{D}_{\nu_{KL}}(P^{(2)} || R), \tag{11}$$

where  $R = \lambda \cdot P^{(1)} + (1 - \lambda) \cdot P^{(2)}$ ,  $\lambda \in [0, 1]$  is some probability distribution from  $\mathcal{P}(X)$ . Obviously,  $\mathcal{D}_{\nu_{JS}}(P^{(1)}, P^{(2)}) = \mathcal{D}_{\nu_{1/2}}(P^{(1)}, P^{(2)})$ . The following relation of the Jensen-Shannon divergence to the Shannon entropy is valid:

$$\mathcal{D}_{\nu_{JS}}(P^{(1)}, P^{(2)}) = h \left( \frac{P^{(1)} + P^{(2)}}{2} \right) - \frac{h(P^{(1)}) + h(P^{(2)})}{2}, \tag{12}$$

where

$$h(P) = - \sum_{i=1}^n P_i \ln P_i \tag{13}$$

is the Shannon entropy,  $P \in \mathcal{P}(X)$ ,  $P = (P_1, P_2, \dots, P_n)$ ,  $0 \leq \mathcal{D}_{\nu_{JS}} \leq \log 2$ .

We could continue the list of definitions of divergence, but the definitions above will be enough to discuss the main problem of the research in this article. It would also be natural to describe the relationships between the divergences mentioned here and the entropy measures. However, this would take us far and go beyond the problems developed in the article.

It should be noted that in recent years, fundamental studies on distance, similarity, and difference relations in the space of finite probability distributions are hardly available. We only find the use of classical definitions in important practical problems. Let us consider some of them: Ref. [13] proposes an algorithm to measure the similarity between nominal variables of the same attribute based on the fact that the similarity between nominal variables depends on the relationship between attributes subsets. This algorithm uses the difference in the distribution which is quantified by f-divergence to form a feature vector

of nominal variables. In [14], neighborhood-based collaborative filtering is considered. A new similarity scheme is proposed that breaks free of the constraint of the co-rated items. Moreover, an item similarity measure based on the Kullback–Leibler divergence is presented, which identifies the relation between items based on the probability distribution of ratings. In [15] the use of a distribution-free overlapping measure as an alternative way to quantify sample differences and assess research hypotheses expressed in terms of Bayesian evidence is illustrated. The main features and potential of the overlapping index by means of three empirical applications are presented. The distance between probability distributions of different dimensions is proposed in [16]. Ref. [17] introduces fuzzy parameters and assesses the similarity between probability distributions using the fuzzy extended Kullback–Leibler divergence. In [18] a new divergence measure is proposed based on the Kullback–Leibler divergence to measure the difference between different basic probability assignments (BPAs) for the Dempster–Shafer belief structure. The book [19] covers the methods of metric distances and their application in probability theory and other fields. The methods are fundamental in the study of limit theorems and generally in assessing the quality of approximations to a given probabilistic model.

### 1.3. On Non-Additive, Monotone Measures and Their Probability Representations in Aggregation Functions in the MADM Environment: A Basic Motivation of the Work

As is known, in stochastic modeling, the construction of the probability distribution in the space of the system states  $X = \{x_1, \dots, x_n\}$  is reduced to the calculations of the frequency distribution of the system states from the prehistoric objective data obtained by conducting random experiments. When studying different populations, additive aggregation functions such as expected value, weighted expected value, and others are mainly used to estimate basic population parameters.

However, today we often encounter such experiments where there is very little or no prehistoric objective data. At such times, the only source of estimation of population parameters is experts and their knowledge. In processing these data, we use non-additive, monotone (we often call them fuzzy) measures instead of additive, probability measures to evaluate the degree of activity of the system state or group of states [20–22]. The syntactic representations of expert assessments on the population are often quite intellectually free. We often use the modern generalizations and extensions of L. Zadeh’s fuzzy set theory developed today to construct the corresponding semantic forms. Therefore, in aggregation procedures, non-additive but monotone aggregation integral functions are used, such as the finite integrals of Choquet [20] and Sugeno [21] and others. By then, in such type aggregation functions two poles—the index of uncertainty of expert assessments (fuzzy measures) and the imprecision characteristic (fuzzy sets, variables)—are considered.

**Definition 1** [22]. Let  $X = \{x_1, \dots, x_n\}$  be some finite set. Denote the algebra of all its subsets by  $2^X$ . We will say that the set function  $g : 2^X \rightarrow [0, 1]$  is a fuzzy measure if

$$(i) \ g(\emptyset) = 0, \quad (ii) \ g(X) = 1, \quad (iii) \ \text{if } A \subset B, \ A, B \in 2^X \text{ then } g(A) \leq g(B). \quad (14)$$

On the basis of the reasoning presented above, it can be concluded that it is essential for aggregating expert assessments in different types of problems, such as face recognition, data classification, decision making, forecasting, and others, to identify the uncertainty index—fuzzy measure; Additionally, on the space of all fuzzy measures  $G(X)$  defined on  $X$ , let us construct the relations of their similarity and difference as it is performed in the corresponding stochastic analysis (see the previous subsection) [21]. However, in this problem, we face an almost unsolvable problem: From the number of values of the fuzzy measure,  $2^{|X|}$  and not  $|X|$  as it is in the case of the probability distribution (measure), it is clear that the direct extensions of the definitions of the distance and divergence between the probabilistic measures given in Sections 1.1 and 1.2 for fuzzy measures are meaningless. As is known, a fuzzy measure is completely equated with its associated probability class.

**Definition 2** [23]. Let us say that  $g : 2^X \rightarrow [0, 1]$  is some fuzzy measure on  $2^X$  and  $S_n$  represents the class of all permutations of the elements of the set  $\{1, 2, \dots, n\}$ . For each permutation  $\sigma = (\sigma(1), \sigma(2), \dots, \sigma(n)) \in S_n$

$$\begin{aligned}
 P_\sigma(x_{\sigma(1)}) &= g(\{x_{\sigma(1)}\}), \\
 P_\sigma(x_{\sigma(2)}) &= g(\{x_{\sigma(1)}, x_{\sigma(2)}\}), \\
 &\dots\dots\dots \\
 P_\sigma(x_{\sigma(i)}) &= g(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}) - g(\{x_{\sigma(1)}, \dots, x_{\sigma(i-1)}\}), \\
 &\dots\dots\dots \\
 P_\sigma(x_{\sigma(n)}) &= 1 - g(\{x_{\sigma(1)}, \dots, x_{\sigma(n-1)}\}), \quad g(\{x_{\sigma(0)}\}) \equiv 0.
 \end{aligned}
 \tag{15}$$

Probability distribution  $P_\sigma = \{P_\sigma(x_{\sigma(1)}), \dots, P_\sigma(x_{\sigma(n)})\}$  is called probability distribution associated with fuzzy measure  $g$ , and  $\{P_\sigma(\cdot)\}_{\sigma \in S_n}$  is called associated probability class (APC).

It should be noted that, due to its non-additivity, the fuzzy measure is a rather inflexible tool when it presents the index of uncertainty in the models of any difficult, complex problem, for example, in MADM aggregation operators. In fact, its values are used in practice in the integral operators of Choquet and Sugeno. Of course, logically, fuzzy measure APC should have more possibilities of use as additive measures, although they form a single class whose “direct insertion” into aggregations requires quite convenient techniques. Fuzzy measure APC has been widely used by authors in their research, especially generalizations of MADM aggregation operators where the fuzzy measure is replaced by probability distributions from its APC. Consider a brief overview of these studies. New fuzzy aggregation operators based on the APC for the extension of the finite Choquet integral are presented in [24]. In [25] the authors used associated probabilities in associated probability intuitionistic fuzzy aggregations for MADM. Probabilistic OWA and other weighted aggregation operators are generalized to fuzzy associated probability operators for fuzzy intuitionistic fuzzy environments. For the same purpose, in [26], APC is used in the construction of new aggregations. Particularly, in interactive MADM under q-rung orthopair fuzzy and q-rung picture linguistic environments. In [26] APC is used in MADM aggregations with interacting attributes for the extension of the Choquet integral under q-rung orthopair fuzzy information. Constructed operators are included in the definition of new—type objective functions in the multi-objective facility location selection problem. In [27] associated probabilities aggregations are developed in multi—stage investment decision-making under q-rung orthopair fuzzy environment. In [28], new associated statistical parameters are developed with the APC aggregations in the Interactive MADM. The APC are developed in the new fuzzy extensions on binomial distribution in [29]. In [30] possibilistic simulation—based interactive fuzzy MADM is constructed on the basis of the APC under discrimination q-rung picture linguistic information. In the same article, associated probabilities are used in the processing of insufficient expert data for the MADM.

As we have already noticed, there are no new studies on the definition of metric, similarity, or divergence relations in fuzzy measures space. Only two works were found [31,32], where the notion of metric for two finite fuzzy measures is introduced. Our study is the first attempt in this direction for finite fuzzy measures. As for the introduction of the mentioned relations in the space of finite probability distributions, this was conducted a long time ago, and a significant place has already been given to this issue in this article too. The new definitions of the mentioned relations in the finite fuzzy measures space, which are based on the APCs of the fuzzy measures, are a kind of generalization of these definitions. It is clear that, in recent years, there have been no new important studies of the mentioned definitions in finite probability spaces. Therefore, what we have included in the article are only the publications published in recent years that deal with the applications of similarity and difference relations of finite probability distributions (Section 1.2).

Based on the above presented, it is already possible to talk about the main motivation of this research. As it was said, in many studies, the use of a finite fuzzy measure in various types of problems is equated with the use of a class of probability distributions associated with it. Definitions of distance, similarity or divergence indexes between finite probability distributions are determined by elementary arithmetic calculations of probability distribution values at the points of phase space. Why should we not use this idea for finite fuzzy measures in the definitions of similar indexes if the probability distribution would be changed by the class of associated probabilities—APC? This task can be formulated formally in the case of distance in general as follows:

As was mentioned, the fuzzy measure  $g$  is uniquely represented by the so-called associated probability class  $APC(g) \subset \mathcal{P}(X)$  [31], and the distance between two fuzzy measures is reduced to the distance between their associated probability classes. If  $g_1$  and  $g_2$  are two fuzzy measures, and  $APC(g_i)$ ,  $i = 1, 2$  are classes of associated probabilities, then in general the distance between them is defined as

$$D(g_1, g_2) = D_{\mathcal{P}}(APC(g_1), APC(g_2)). \quad (16)$$

We will use this idea in the current study to generalize classical measures of similarity and divergence between two fuzzy measures from analogs of probabilistic measures, respectively.

If this could be done, then many interesting practical and applied problems, whose uncertainty index researchers consider the fuzzy measure, would be reduced to algebraic manipulations of the elements of the class of associated probabilities. After all, this class is a very convenient tool, because each of its elements is a finite probability distribution. All this would make it easier to solve many problems, because we will not be dealing with a monotone, but an additive class of measures. Let us cite some examples of such problems: effective solution of fuzzy measure identification, which is one of the problems of this research; determining the difference between the Dempster—Shafer belief structures; measuring distribution similarities between samples; problems of determining the similarity of non-additive measures of different dimensions; problems of constructing user similarity relations in collaborative filtering models; metrics introduction problems in the phase space of non-additive and interacting parameters of machine learning; clustering and classification of complex objects and other problems.

The second section presents the preliminary concepts. In particular, certain classes of fuzzy measures are briefly described: the Sugeno  $\lambda$ -additive measures [21] and Choquet second—order capacities equivalent to supermodular and submodular dual fuzzy measures [20,33,34]. Representations of their associated probabilities, as well as Dempster—Shafer’s belief structure [35,36] and its probabilistic representations, are formulated. The third section discusses divergence and similarity indexes, as well as new definitions of distance for two fuzzy measures. The proof of the correctness of the generalizations of the definitions of distance, divergence, and similarity measures is presented in the fourth section. The use of these generalized concepts in fuzzy measure identification problems for different fuzzy measure classes is discussed in the fifth section. The obtained results and prospects for research development are presented in the sixth section.

## 2. Preliminary Concepts

### 2.1. Sugeno $\lambda$ -Additive Measures and the Choquet Second-Order Capacities: Representations of Their Associated Probabilities

In this subsection, we briefly introduce the important and frequently used fuzzy measure classes in practice. Non-additive monotone measures (named  $\lambda$ -additive fuzzy measure) were first introduced by Sugeno [21]. He called monotone measures fuzzy measures, and later A. Kandel called them fuzzy statistics [37].

**Definition 3 [21].** Let  $X = \{x_1, \dots, x_n\}$  be some finite set. Denote the algebra of all its subsets by  $2^X$ . A fuzzy measure  $g_\lambda : 2^X \rightarrow [0, 1]$  ( $\lambda > -1$ ) is called an  $\lambda$ -additive fuzzy measure if for every two sets  $A, B \in 2^X$ ,  $A \cap B = \emptyset$

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda \cdot g_\lambda(A) \cdot g_\lambda(B), \tag{17}$$

by the following normalization condition

$$\frac{1}{\lambda} \left\{ \prod_{x_i \in X} (1 + \lambda g_\lambda(\{x_i\})) - 1 \right\} = 1. \tag{18}$$

It is easy to show that  $\forall A \in 2^X$

$$g_\lambda(A) = \frac{1}{\lambda} \left\{ \prod_{x_i \in A} (1 + \lambda g_\lambda(\{x_i\})) - 1 \right\}. \tag{19}$$

It is not difficult to write the associated probabilities of  $\lambda$ -additive fuzzy measure:

$$\forall \sigma \in S_n, \quad P_\sigma(x_i) = g_\lambda(\{x_i\}) \cdot \prod_{j=1}^{i(\sigma)-1} (1 + \lambda g_\lambda(\{x_{\sigma(j)}\})), \tag{20}$$

where  $i = 1, 2, \dots, n$ ;  $i(\sigma)$  is the index of location of element  $x_i$  in the permutation

$$\sigma = (\sigma(1), \dots, \sigma(n)). \text{ If } i(\sigma) = 1, \text{ then } \prod_{j=1}^0 \equiv 1.$$

**Definition 4 [23].**  $g_*$  and  $g^*$  dual fuzzy measures on  $2^X$  are called the Choquet second-order lower and upper capacities, respectively, if  $\forall A, B \in 2^X$

$$\begin{aligned} g_*(A \cap B) + g_*(A \cup B) &\geq g_*(A) + g_*(B) \\ g^*(A \cap B) + g^*(A \cup B) &\leq g^*(A) + g^*(B) \end{aligned} \tag{21}$$

where  $g^*(A) = 1 - g_*(\bar{A})$ .

The Choquet second-order capacities are a fairly broad class of fuzzy measures. Sugeno  $\lambda$ -additive fuzzy measures also belong to them. It is not difficult to show that for  $g_{-1/(1+\lambda)}^*$  is dual to  $g_{\lambda^*}$ .

It is easy to prove that

**Proposition 1 [23].** If  $g_*, g^* : 2^X \rightarrow [0, 1]$  are dual fuzzy measures, then they have a common associated probability class  $\{P_\sigma(\cdot)\}_{\sigma \in S_n}$  and for every  $\sigma \in S_n$  we have  $P_\sigma^{(g^*)}(\cdot) = P_{\sigma^*}^{(g_*)}(\cdot)$ , where  $\sigma^*$  is a dual permutation to the permutation  $\sigma$ :

$$\sigma(i) = \sigma^*(n - i + 1), \quad i = 1, \dots, n. \tag{22}$$

Let us introduce the following notations:

Suppose  $\mathcal{G}(X)$  is a class of all fuzzy measures defined on  $2^X$  and  $\mathcal{G}_{Ch}^{(2)}(X)$  is the Choquet second-order capacities class on  $2^X$  ( $\mathcal{G}_{Ch}^{(2)}(X) \subset \mathcal{G}(X)$ ); Let  $\mathcal{G}_S(X)$  be Sugeno  $\lambda$ -additive fuzzy measures class on  $2^X$  ( $\mathcal{G}_S(X) \subset \mathcal{G}_{Ch}^{(2)}(X) \subset \mathcal{G}(X)$ ). It is clear that if  $P : 2^X \rightarrow [0, 1]$  is any probabilistic measure (probability distribution), then it can be formally considered as a fuzzy measure and its associated probability class consists of one element and it coincides

with this distribution itself. Let us denote by  $\mathcal{P}(X)$  the class of all probabilistic measures defined on  $2^X$ , then

$$\mathcal{P}(X) \subset \mathcal{G}_S(X) \subset \mathcal{G}_{Ch}^{(2)}(X) \subset \mathcal{G}(X). \tag{23}$$

**Proposition 2 [23].** *If  $g_*, g^* : 2^X \rightarrow [0, 1]$  are the Choquet dual capacities of the second order ( $g_*, g^* \in \mathcal{G}_{Ch}^{(2)}(X)$ ), then  $\forall A \subseteq X$ :*

$$g_*(A) = \min_{\sigma \in S_n} P_\sigma(A); \quad g^*(A) = \max_{\sigma \in S_n} P_\sigma(A), \tag{24}$$

where  $\{P_\sigma(\cdot)\}_{\sigma \in S_n}$  is the common associated probabilities class of dual measures  $g_*$  and  $g^*$ .

### 2.2. Dempster-Shafer’s Belief Structure and Its Probabilistic Representations

The theory of evidence is based on two dual fuzzy measures—belief (*Bel*) and plausibility (*Pl*) measures [35,36]. It is easy to show that these two classes of dual measures are also subclasses of second-order Choquet capacities. These two fuzzy measures on  $2^X$  are uniquely determined by the so-called basic probability assignment on  $2^X$  (*BPA*) [36]:

$$m : 2^X \rightarrow [0, 1] \tag{25}$$

with the following two conditions:

$$(i) \ m(\emptyset) = 0; \quad (ii) \ \sum_{B \in 2^X} m(B) = 1. \tag{26}$$

Of course, generally speaking,  $m$  is probability distribution on  $2^X$ , not on  $X$ ! Then

$$\begin{aligned} Pl(A) &= \sum_{B: A \cap B \neq \emptyset} m(B), \quad \forall A \in 2^X, \\ Bel(A) &= \sum_{B: B \subset A} m(B), \quad \forall A \in 2^X. \end{aligned} \tag{27}$$

There is also feedback:

$$m(A) = \sum_{B: B \subset A} (-1)^{|A \setminus B|} Bel(B), \quad \forall A \in 2^X. \tag{28}$$

Each set  $A \in 2^X$ , for which  $m(A) > 0$ , is called a focal element. The pair  $\langle \mathcal{F}, m \rangle$  is called the body of evidence, where

$$\mathcal{F} = \{A, A \in 2^X, m(A) > 0\}. \tag{29}$$

It is easy to show that

$$m(A) = \sum_{B \in \mathcal{F}; B \subset A} (-1)^{|A \setminus B|} \min_{\sigma \in S_n} P_\sigma^{(Bel)}(B), \quad \forall A \in 2^X, \tag{30}$$

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

**Definition 5 [38].** *A possibility measure  $Pos$  on  $2^X$  is called a fuzzy measure for which there is a possibility distribution on  $X$ ,  $\pi : X \rightarrow [0, 1)$ ,  $\pi(x_i) = Pos(\{x_i\})$ ,  $i = 1, \dots, n$ ;  $\exists x_0 : \pi(x_0) \equiv 1$ , such that*

$$Pos(A) = \max_{x_i \in A} \pi(x_i), \quad \forall A \in 2^X. \tag{31}$$

The necessity measure  $Nes$  is dual to the possibility measure, so that  $Nes(A) = 1 - Pos(\bar{A})$ ,  $\forall A \in 2^X$  [38]. It is known that for every  $Pos$  measure on  $2^X$  there is a so-called consonant data body [36]

$$\mathcal{F} = \{A_{l_1} \subset A_{l_2} \subset \dots \subset A_{l_k}\}, \quad A_{l_i} = \{x_1, x_2, \dots, x_{l_i}\}, \quad i = 1, 2, \dots, k, \quad (32)$$

and by the possibility distribution  $m_{l_i} \equiv m(A_{l_i})$ ,  $i = 1, \dots, k$ ;  $\pi_j = \pi(x_j)$ ,  $j = 1, \dots, n$ , there is the following connection between them:

$$\begin{cases} \pi_i = \sum_{j: x_i \in A_{l_j}} m_{l_j}, \quad i = 1, 2, \dots, n \\ m_{l_j} = \pi_{l_j} - \pi_{l_{j+1}}, \quad \pi_{l_{k+1}} \equiv 0, \quad j = 1, 2, \dots, k. \end{cases} \quad (33)$$

Let  $\{P_\sigma^{(Pos)}(\cdot)\}_{\sigma \in S_n}$  be the associated probabilities class of the  $Pos$  measure. Then the connections between elements  $\{\pi_i\}$ ,  $\{m_{l_i}\}$  and  $\{P_\sigma^{(Pos)}(\cdot)\}_{\sigma \in S_n}$  are easily obtained:

$$P_\sigma^{(Pos)}(\{x_{\sigma(i)}\}) = \max_{l=1, \dots, i} \{\pi_{\sigma(l)}\} - \max_{l=1, \dots, i-1} \{\pi_{\sigma(l)}\} = \max_{j=1, \dots, i: x_{\sigma(j)} \in A_{l_j}} \sum m_{l_j} - \max_{j=1, \dots, i-1: x_{\sigma(j)} \in A_{l_j}} \sum m_{l_j}. \quad (34)$$

On the other hand

$$\pi_i = Pos(\{x_i\}) = \max_{\sigma \in S_n} P_\sigma^{(Pos)}(\{x_i\}), \quad i = 1, 2, \dots, n. \quad (35)$$

Since  $Pos$  is the second-order Choquet capacity, we obtain

$$m_{l_j} = \pi_{l_j} - \pi_{l_{j+1}} = \max_{\sigma \in S_n} P_\sigma^{(Pos)}(x_{l_j}) - \max_{\sigma \in S_n} P_\sigma^{(Pos)}(x_{l_{j+1}}), \quad j = 1, \dots, k. \quad (36)$$

Denote by  $\mathcal{G}_{Pos}(X)$  the class of all possibility measures defined on  $2^X$ .

Let  $m$  be some  $BPA\langle \mathcal{F}, m \rangle$ ,  $\mathcal{F} = \{A_1, A_2, \dots, A_k\}$ . Suppose for each focal element  $A_j$ ,  $j = 1, \dots, k$  we have  $|A_j|$ -dimensional  $W_j$  weight vector  $W_j = \langle w_j(1), \dots, w_j(|A_j|) \rangle$  in this focal element for each  $x \in X$ , such that  $w_j(i) \in [0, 1]$ ,  $\sum_{i=1}^{|A_j|} w_j(i) = 1$ . This is called the allocation vector of the element  $A_j$ . R. Yager [39] has studied the set function

$$g(A) = \sum_{j=1}^k [m(A_j) \cdot \sum_{i=1}^{|A_j \cap A|} w_j(i)], \quad (37)$$

which represents the fuzzy measure associated with the belief structure  $\langle \mathcal{F}, m \rangle$  on  $2^X$ . Denote by  $\mathcal{G}_{\langle \mathcal{F}, m \rangle}(X)$  is the class of fuzzy measures associated with the body of evidence  $\langle \mathcal{F}, m \rangle$  on  $2^X$ . Obviously, each of such fuzzy measures is uniquely defined by the class of certain allocation vectors  $W = \{W_1, \dots, W_k\}$ . If  $\forall j = 1, \dots, k$ ,  $W_j$  is chosen so that  $w_j(1) = 0$  and  $w_j(i) = 0$ ,  $i = 2, \dots, |A_j|$ , then the fuzzy measure  $g$  coincides with the measure  $Pos$ . And vice versa, if  $w_j(|A_j|) = 1$  and  $w_j(i) = 0$ ,  $i = 1, \dots, |A_j| - 1$ , then  $g$  coincides with the measure  $Nes$ . Let us note here that all fuzzy measures of this class have a common Shapley entropy [39]. It is easy to see the relationship between the associated fuzzy measure, the associated probability class, the focal elements, and the allocation vector class  $W = \{w_1, \dots, w_k\}$ .

$$\begin{aligned} P_\sigma^{(g)}(\{x_{\sigma(i)}\}) &= \sum_{j=1}^k m(A_j) \cdot \left[ \sum_{l=1}^{|A_j \cap \{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}|} w_{j(l)} - \sum_{l=1}^{|A_j \cap \{x_{\sigma(1)}, \dots, x_{\sigma(i-1)}\}|} w_{j(l)} \right] \\ &= \sum_{A_j \in \mathcal{F}: A_j \cap \{x_{\sigma(i)}\} \neq \emptyset} m(A_j) w_j(|A_j \cap \{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}|). \end{aligned} \quad (38)$$

### 3. Divergence, Similarity and Distance Parameters for Two Fuzzy Measures

In this section, we will realize the main problem formulated in the introduction, which is formulated in the case of distance in the form (16). We will try to generalize the distance and divergence measures of two probability distributions to the case of two fuzzy measures. As it was shown in the previous paragraph, any fuzzy measure  $g \in \mathcal{G}(X)$  generates its APC- $\{P_\sigma^{(g)}(\cdot)\}_{\sigma \in S_n}$ . The reverse statement is also valid.

**Proposition 3 [23].** *Let us be given a fuzzy measure  $g \in \mathcal{G}(X)$  and its APC- $\{P_\sigma^{(g)}(\cdot)\}_{\sigma \in S_n}$ . If  $A = \{x_{i_1}, x_{i_2}, \dots, x_{i_s}\} \in 2^X$  is some set, then*

$$g(A) = P_\tau^{(g)}(A), \tag{39}$$

where in permutation  $\tau = (\tau(1), \tau(2), \dots, \tau(s), \dots, \tau(n))$  of the first  $s$  indexes coincide with indexes  $\langle i_1, i_2, \dots, i_s \rangle$ :  $\tau(j) = i_j, j = 1, 2, \dots, s$ .

Therefore, with associated probabilities, we always get the  $g$  fuzzy measure's values on the arguments  $A \in 2^X$ . Thus, we will say that the fuzzy measure  $g \in \mathcal{G}(X)$  is given if and only if the class of its associated probabilities  $\{P_\sigma^{(g)}(\cdot)\}_{\sigma \in S_n}$  is known.

The main concept of our study is to relate the measurement of similarity or divergence between two fuzzy measures to the similarity and difference between their APCs (15).

#### 3.1. Distance Generalizations for Two Fuzzy Measures on $\mathcal{G}(X)$

**Definition 6.** *Let  $T^m \equiv \{(z_1, z_2, \dots, z_m) \in R^m \mid z_i \geq 0, i = 1, 2, \dots, m\}$ . Let also  $F : T^m \rightarrow R^+$  be a function. We say that  $F$  is a distance generating function if it satisfies the following five properties:*

- (1)  $F(z_1, z_2, \dots, z_m) = 0$  if  $z_1 = z_2 = \dots = z_m = 0$
- (2) If  $z'_i \leq z''_i$  for any  $i$  then  $F(z'_1, z'_2, \dots, z'_m) \leq F(z''_1, z''_2, \dots, z''_m)$

That is,  $F$  is monotonically non-decreasing.

$$(3) F(z'_1 + z''_1, z'_2 + z''_2, \dots, z'_m + z''_m) \leq F(z'_1, z'_2, \dots, z'_m) + F(z''_1, z''_2, \dots, z''_m). \tag{40}$$

That is,  $F$  is sub-additive.

$$(4) F(z, z, \dots, z) = z.$$

That is,  $F$  is idempotent

$$(5) F(z_{\sigma(1)}, z_{\sigma(2)}, \dots, z_{\sigma(m)}) = F(z_1, z_2, \dots, z_m), \quad \forall \sigma \in S_m.$$

That is,  $F$  is symmetric.

Sort the permutations  $n! = m \quad \sigma = (\sigma(1), \dots, \sigma(n))$  from  $S_n$  by some criterion to renumber them from 1 to  $m$ , and, therefore, we get the associated probability class  $\{P_\sigma(\cdot)\}_{\sigma \in S_n}$  as a  $m$ -dimensional vector  $(P_1, P_2, \dots, P_m)$  where  $m = n!$ . Let  $d$  be a distance on  $\mathcal{P}(X)$ .

**Definition 7.** *Let us be given two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$ . A binary argument function*

$$\mathcal{D}^{(F,d)}(g_1, g_2) = F(d(P_1^{(g_1)}, P_1^{(g_2)}), \dots, d(P_m^{(g_1)}, P_m^{(g_2)})) \tag{41}$$

is called the distance between two fuzzy measures, where  $(P_1^{(g_1)}, \dots, P_m^{(g_1)})$  and  $(P_1^{(g_2)}, \dots, P_m^{(g_2)})$  are associated probability classes of fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$ , respectively.

It is not difficult to show that the function  $\mathcal{D}^{(F,d)} : \mathcal{G}(X) \times \mathcal{G}(X)$  defined by the distance generator function has distance properties. Here are two examples of generator function  $F$  that we will use in the future:

$$F_q(z_1, z_2, \dots, z_m) \equiv \left( \frac{1}{m} \sum_{i=1}^m z_i^q \right)^{1/q}, \quad q \geq 1$$

$$F_{\max}(z_1, z_2, \dots, z_m) = \max_{1 \leq i \leq m} \{z_i\}. \tag{42}$$

If the distance  $d$  between probability distributions is calculated by the following two formulas:

$$d_{\max}(P^{(g_1)}, P^{(g_2)}) = \max_{1 \leq i \leq n} |P^{(g_1)}(x_i) - P^{(g_2)}(x_i)|$$

$$d_q(P^{(g_1)}, P^{(g_2)}) = \left( \sum_{i=1}^n |P^{(g_1)}(x_i) - P^{(g_2)}(x_i)|^q \right)^{1/q}, \quad q \geq 1, \tag{43}$$

then it is easy to show.

**Proposition 4.** *Let us be given two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$ . The distances between fuzzy measures and their corresponding dual fuzzy measures coincide*

$$\mathcal{D}^{(F,d)}(g_1, g_2) = \mathcal{D}^{(F,d)}(g_1^*, g_2^*), \tag{44}$$

where  $g_1^*$  and  $g_2^*$  are dual fuzzy measures to  $g_1$  and  $g_2$ , respectively.

As a concrete example, consider the case when

$$F_2(z_1, \dots, z_m) = \sqrt{\frac{1}{m} \sum_{i=1}^m z_i^2} \tag{45}$$

and distance on  $\mathcal{P}(X)$  is defined by the formula:

$$d_2(P^{(1)}, P^{(2)}) = \sqrt{\sum_{i=1}^n (P^{(1)}(x_i) - P^{(2)}(x_i))^2}, \tag{46}$$

then

$$\mathcal{D}^{(F_2, d_2)}(g_1, g_2) = \sqrt{\frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n (P_j^{(g_1)}(x_i) - P_j^{(g_2)}(x_i))^2} =$$

$$\sqrt{\frac{1}{n!} \sum_{\sigma \in S_n} \sum_{i=1}^n (P_{\sigma(i)}^{(g_1)}(x_{\sigma(i)}) - P_{\sigma(i)}^{(g_2)}(x_{\sigma(i)}))^2}. \tag{47}$$

Let us introduce the concepts of associated parameters with associated probabilities of any fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$ , which will be analogous to the definitions provided in the introduction of the article. Let  $\{P_{\sigma}^{(g_1)}\}_{\sigma \in S_n}$  and  $\{P_{\sigma}^{(g_2)}\}_{\sigma \in S_n}$  be the APCs for the fuzzy measures  $g_1$  and  $g_2$  from  $\mathcal{G}(X)$ , respectively.

**Definition 8.** 1. *Bhattacharyya’s F—coefficient for two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$  is called*

$$BC^{(F)}(g_1, g_2) = F(BC(P_1^{(g_1)}, P_1^{(g_2)}), \dots, BC(P_m^{(g_1)}, P_m^{(g_2)})), \tag{48}$$

where  $BC(P_i^{(g_1)}, P_i^{(g_2)}) = \sum_{j=1}^n [P_i^{(g_1)}(x_j), P_i^{(g_2)}(x_j)]^{1/2}, \quad i = 1, \dots, m$

2. *Then Bhattacharyya’s F—distance between two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$  is called*

$$\mathcal{D}_B^{(F)}(g_1, g_2) = F(-\log(BC(P_1^{(g_1)}, P_1^{(g_2)})), \dots, -\log(BC(P_m^{(g_1)}, P_m^{(g_2)}))). \tag{49}$$

Other generalized  $F$ —distances between two fuzzy measures are defined similarly. Calculate the distance for the  $F = F_{max}$  function:

$$\begin{aligned} \mathcal{D}_B^{(F_{max})}(g_1, g_2) &= \max_{1 \leq i \leq m} \left\{ -\log(BC(P_i^{(g_1)}, P_i^{(g_2)})) \right\} = -\min_{\sigma \in S_n} \left\{ \log(BC(P_\sigma^{(g_1)}, P_\sigma^{(g_2)})) \right\} \\ &= -\min_{\sigma \in S_n} \left\{ \log \left( \sum_{i=1}^n [(P_\sigma^{(g_1)}(x_{\sigma(i)}) \cdot P_\sigma^{(g_2)}(x_{\sigma(i)})]^{1/2} \right) \right\}, \end{aligned} \tag{50}$$

while for  $F = F_q, (q \geq 1)$ , we have

$$\begin{aligned} \mathcal{D}_B^{(F_q)}(g_1, g_2) &= \sqrt[q]{\frac{1}{m} \sum_{i=1}^m [\log \frac{1}{BC(P_i^{(g_1)}, P_i^{(g_2)})}]^q} = \\ &= \sqrt[q]{\frac{1}{n!} \sum_{\sigma \in S_n} [\log \frac{1}{\sum_{i=1}^n (P_\sigma^{(g_1)}(x_{\sigma(i)}) P_\sigma^{(g_2)}(x_{\sigma(i)}))^{1/2}}]^q}. \end{aligned} \tag{51}$$

For the particular case  $F = F_1$ , we have simple expression

$$\begin{aligned} \mathcal{D}_B^{(F_1)}(g_1, g_2) &= -\frac{1}{m} \sum_{\sigma \in S_n} \log(BC(P_\sigma^{(g_1)}, P_\sigma^{(g_2)})) = \\ &= -\frac{1}{n!} \log \left\{ \left[ \prod_{\sigma \in S_n} \left( \sum_{i=1}^n (P_\sigma^{(g_1)}(x_{\sigma(i)}) \cdot P_\sigma^{(g_2)}(x_{\sigma(i)}))^{1/2} \right) \right] \right\}. \end{aligned} \tag{52}$$

In the same way, we can construct Bhattacharyya’s  $F$ —angle between two fuzzy measures. We also similarly can construct the Helling  $F$ —distance for two fuzzy measures (omitted here).

### 3.2. Divergence Generalizations and Similarity Index for Two Fuzzy Measures

Now let us turn to generalizations of divergences for two fuzzy measures.

**Definition 9.** Rényi  $F$ —divergence between two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$  is called:

$$\mathcal{D}_{\mathcal{R}}^{(F, \alpha)}(g_1 \parallel g_2) = F(\mathcal{D}_{\mathcal{R}}^{(\alpha)}(P_1^{(g_1)} \parallel P_1^{(g_2)}), \dots, \mathcal{D}_{\mathcal{R}}^{(\alpha)}(P_m^{(g_1)} \parallel P_m^{(g_2)})), \tag{53}$$

where  $0 < \alpha < \infty, \alpha \neq 1$ .

For the function  $F = F_q$  we obtain

$$\begin{aligned} \mathcal{D}_{\mathcal{R}}^{(F_q, \alpha)}(g_1 \parallel g_2) &= \sqrt[q]{\frac{1}{m} \sum_{\sigma \in S_n} [\mathcal{D}_{\mathcal{R}}^{(\alpha)}(P_\sigma^{(g_1)}, P_\sigma^{(g_2)})]^q} = \\ &= \sqrt[q]{\frac{1}{m(\alpha - 1)^q} \sum_{\sigma \in S_n} \left( \log \left( \sum_{i=1}^n \frac{(P_\sigma^{(g_1)}(x_{\sigma(i)}))^\alpha}{(P_\sigma^{(g_2)}(x_{\sigma(i)}))^{\alpha-1}} \right) \right)^q}. \end{aligned} \tag{54}$$

**Definition 10.** The Kullback—Leibler  $F$ —divergence between two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$  is called:

$$\mathcal{D}_{\mathcal{KL}}^{(F)}(g_1 \parallel g_2) = F[\mathcal{D}_{\mathcal{KL}}(P_1^{(g_1)} \parallel P_1^{(g_2)}), \dots, \mathcal{D}_{\mathcal{KL}}(P_m^{(g_1)} \parallel P_m^{(g_2)})]. \tag{55}$$

As an example, let us build the Kullback–Leibler  $F_q$ –divergence ( $\alpha \rightarrow 1$ ).

$$\mathcal{D}u_{KL}^{(F_q)}(g_1 \parallel g_2) = \sqrt[q]{\frac{1}{m} \sum_{\sigma \in S_n} [\mathcal{D}u_{KL}(P_\sigma^{(g_1)}, P_\sigma^{(g_2)})]^q} = \sqrt[q]{\frac{1}{n!} \sum_{\sigma \in S_n} \left( \sum_{i=1}^n (P_\sigma^{(g_1)}(x_{\sigma(i)}) \cdot \log \frac{P_\sigma^{(g_1)}(x_{\sigma(i)})}{P_\sigma^{(g_2)}(x_{\sigma(i)})} \right)^q}.$$

Let us also construct the Kullback–Leibler  $F_{\max}$ –divergence.

$$\mathcal{D}u_{KL}^{(F_{\max})}(g_1 \parallel g_2) = \max_{\sigma \in S_n} \left\{ \mathcal{D}u_{KL}(P_\sigma^{(g_1)}, P_\sigma^{(g_2)}) \right\} = \max_{\sigma \in S_n} \left\{ \sum_{i=1}^n (P_\sigma^{(g_1)}(x_{\sigma(i)}) \cdot \log \frac{P_\sigma^{(g_1)}(x_{\sigma(i)})}{P_\sigma^{(g_2)}(x_{\sigma(i)})} \right\}. \tag{56}$$

**Definition 11.** The Jeffrey’s  $F$ –divergence for two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$  is called:

$$\mathcal{D}u_J^{(F)}(g_1, g_2) \triangleq \mathcal{D}u_{KL}^{(F)}(g_1 \parallel g_2) + \mathcal{D}u_{KL}^{(F)}(g_2 \parallel g_1) = \mathcal{D}u_J^{(F)}(g_2, g_1). \tag{57}$$

Let us construct the Jeffrey’s divergence for  $F = F_q$ :

$$\mathcal{D}u_J^{(F_q)}(g_1, g_2) = \sqrt[q]{\frac{1}{n!} \sum_{\sigma \in S_n} \left( \sum_{i=1}^n [(P_\sigma^{(g_1)}(x_{\sigma(i)}) - P_\sigma^{(g_2)}(x_{\sigma(i)})] \cdot \log \frac{P_\sigma^{(g_1)}(x_{\sigma(i)})}{P_\sigma^{(g_2)}(x_{\sigma(i)})} \right)^q}. \tag{58}$$

For  $F = F_{\max}$  we obtain:

$$\mathcal{D}u_J^{(F_{\max})}(g_1, g_2) = \max_{\sigma \in S_n} \left[ \sum_{i=1}^n (P_\sigma^{(g_1)}(x_{\sigma(i)}) - P_\sigma^{(g_2)}(x_{\sigma(i)})) \cdot \log \frac{P_\sigma^{(g_1)}(x_{\sigma(i)})}{P_\sigma^{(g_2)}(x_{\sigma(i)})} \right]. \tag{59}$$

**Definition 12.**  $\langle F, \lambda \rangle$ –divergence for two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$  is called

$$\mathcal{D}u^{(F, \lambda)}(g_1, g_2) \triangleq \lambda \mathcal{D}u_{KL}^{(F)}(g_1 \parallel R) + (1 - \lambda) \mathcal{D}u_{KL}^{(F)}(g_2 \parallel R), \tag{60}$$

where the associated probabilities of a fuzzy measure  $R$  are obtained as a convex combination of the associated probabilities of fuzzy measures  $g_1$  and  $g_2$ .

$$P_\sigma^{(R)}(x_{\sigma(i)}) = \lambda P_\sigma^{(g_1)}(x_{\sigma(i)}) + (1 - \lambda) P_\sigma^{(g_2)}(x_{\sigma(i)}), \quad 0 \leq \lambda \leq 1. \tag{61}$$

If in this definition  $\lambda = 1/2$ , then we get a generalization of Jensen-Shannon’s classical divergence with respect to fuzzy measures.

**Definition 13.** The Jensen-Shannon  $F$ –divergence for two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$  is called

$$\mathcal{D}u_{JS}^{(F)}(g_1, g_2) = \mathcal{D}u^{(F, 1/2)}(g_1, g_2) = \frac{1}{2} \left( \mathcal{D}u_{KL}^{(F)}(g_1 \parallel R) + \mathcal{D}u_{KL}^{(F)}(g_2 \parallel R) \right). \tag{62}$$

Let us construct the Jensen-Shannon  $F = F_q$ –divergence

$$\begin{aligned}
 \mathcal{D}\nu_{JS}^{(F_q)}(g_1, g_2) &= \sqrt[q]{\frac{1}{2^q m} \sum_{i=1}^m \left( \sum_{j=1}^n \left[ P_j^{(g_1)} \cdot \log\left(\frac{2P_j^{(g_1)}}{P_j^{(g_1)} + P_j^{(g_2)}}\right) + P_j^{(g_2)} \cdot \log\left(\frac{2P_j^{(g_2)}}{P_j^{(g_1)} + P_j^{(g_2)}}\right) \right] \right)^q} \\
 &= \sqrt[q]{\frac{1}{2^q n!} \sum_{\sigma \in S_n} \left( \sum_{j=1}^n \left[ P_j^{(g_1)}(x_{\sigma(j)}) \cdot \log\left(\frac{2P_{\sigma}^{(g_1)}(x_{\sigma(j)})}{P_{\sigma}^{(g_1)}(x_{\sigma(j)}) + P_{\sigma}^{(g_2)}(x_{\sigma(j)})}\right) + P_{\sigma}^{(g_2)}(x_{\sigma(j)}) \cdot \log\left(\frac{2P_{\sigma}^{(g_2)}(x_{\sigma(j)})}{P_{\sigma}^{(g_1)}(x_{\sigma(j)}) + P_{\sigma}^{(g_2)}(x_{\sigma(j)})}\right) \right] \right)^q}. \tag{63}
 \end{aligned}$$

Let us also construct the Jensen-Shannon  $F = F_{\max}$ —divergence

$$\mathcal{D}\nu_{JS}^{(F_{\max})}(g_1, g_2) = \frac{1}{2} \max_{\sigma \in S_n} \left\{ \begin{aligned} &\sum_{j=1}^n \left( P_j^{(g_1)}(x_{\sigma(j)}) \cdot \log\left(\frac{2P_{\sigma}^{(g_1)}(x_{\sigma(j)})}{P_{\sigma}^{(g_1)}(x_{\sigma(j)}) + P_{\sigma}^{(g_2)}(x_{\sigma(j)})}\right) + \right. \\ &\left. P_{\sigma}^{(g_2)}(x_{\sigma(j)}) \cdot \log\left(\frac{2P_{\sigma}^{(g_2)}(x_{\sigma(j)})}{P_{\sigma}^{(g_1)}(x_{\sigma(j)}) + P_{\sigma}^{(g_2)}(x_{\sigma(j)})}\right) \right) \end{aligned} \right\}. \tag{64}$$

As we can see, there are many options and combinations of generalizations, however, their further definitions would take us too far from the main topic of the article.

### 3.3. Similarity Index Between Two Fuzzy Measures

In this subsection, we will consider the generalization of the definition of similarity relation to the space of fuzzy measures. We will also provide a short analysis of it.

Huang [40] defined the index of similarity of two probability distributions.

**Definition 14 [40].** If two probability distributions  $P^{(1)}, P^{(2)} \in P(X)$  are defined on  $X$ , then their distributions similarity index is defined as:

$$DSI(P^{(1)}, P^{(2)}) = \frac{\beta(P^{(1)}, P^{(2)})}{\frac{1}{2}[\beta(P^{(1)}) + \beta(P^{(2)})]}, \tag{65}$$

where the expression

$$\beta(P) = \sum_{i=1}^n (P(x_i))^2 = \mathcal{E}_P(P) \tag{66}$$

is called an informity of the distribution  $P \in P(X)$ , and the expression

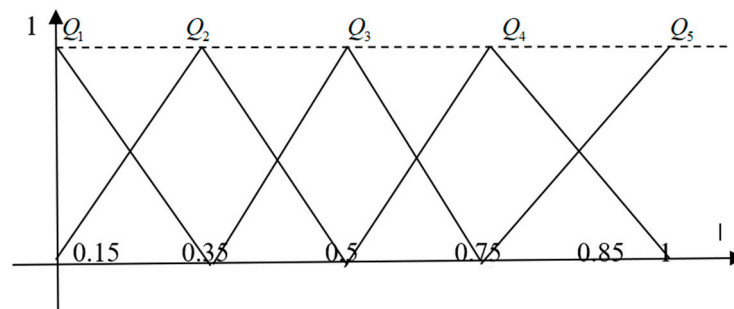
$$\beta(P^{(1)}, P^{(2)}) = \sum_{i=1}^n P^{(1)}(x_i)P^{(2)}(x_i) \tag{67}$$

is called a cross informity of two distributions.

Huang noted [40] that if  $DSI = 0.75, 0.5$ , and  $0.25$ , then the similarity of distributions  $P^{(1)}$  and  $P^{(2)}$  is high, medium and low, respectively. This informational analysis and related synthesis issues are indeed questionable, and many authors have even noted the lesser correctness of this analysis. Perhaps it would be useful to describe the linguistic variable of similarity with fuzzy terms, in order to use it further in synthesis issues. See one of the examples of this in Figure 1:

Obviously, if we start the analysis of the divergences defined above, we have to introduce new strong semantic forms to solve their practical problems of synthesis.

We can easily generalize Huang’s similarity index in the case of two fuzzy measures as well.



**Figure 1.** Fuzzy terms of similarity of two fuzzy measures.  $Q_1$  = “very low similarity”,  $Q_2$  = “low similarity”,  $Q_3$  = “medium similarity”,  $Q_4$  = “high similarity”,  $Q_5$  = “very high similarity”.

**Definition 15.** Huang’s  $F$ —similarity index for two fuzzy measures  $g_1, g_2 \in \mathcal{G}(X)$  is called:

$$DSI^{(F)}(g_1, g_2) = F(DSI(P_1^{(g_1)}, P_1^{(g_2)}), \dots, DSI(P_m^{(g_1)}, P_m^{(g_2)})), \quad m = n!. \quad (68)$$

For  $F = F_q$ —similarity index we have:

$$DSI^{(F_q)}(g_1, g_2) = 2^q \sqrt[q]{\frac{1}{n!} \left[ \sum_{\sigma \in S_n} \left( \frac{\sum_{j=1}^n P_{\sigma}^{(g_1)}(x_j) P_{\sigma}^{(g_2)}(x_j)}{\sum_{i=1}^n [(P_{\sigma}^{(g_1)}(x_j))^2 + (P_{\sigma}^{(g_2)}(x_j))^2]} \right)^q \right]}, \quad (69)$$

and for the  $F = F_{\max}$  case:

$$DSI^{(F_{\max})}(g_1, g_2) = 2 \max_{\sigma \in S_n} \left\{ \frac{\sum_{j=1}^n P_{\sigma}^{(g_1)}(x_j) P_{\sigma}^{(g_2)}(x_j)}{\sum_{i=1}^n [(P_{\sigma}^{(g_1)}(x_j))^2 + (P_{\sigma}^{(g_2)}(x_j))^2]} \right\}. \quad (70)$$

It is easy to show that  $DSI^{(F)}(\cdot) \in [0, 1]$ .

### 3.4. On the Correctness of the Generalizations of Definitions of Distance, Divergence and Similarity Measures

As was mentioned, the APC— $\{P_{\sigma}\}_{\sigma \in S_n}$  of a probability distribution consists of one element, and this element is the probability distribution itself. Obviously, if we consider the definitions developed in the article for two fuzzy measures in relation to distance, divergence and similarity, when these two fuzzy measures are probability ones, then the generalized expressions should coincide with the corresponding classical formulas for the probability measures.

**Proposition 5.** *If in the role of fuzzy measures  $g_1$  and  $g_2$  in Formulas (47)–(64) and (68) we consider probability distributions  $g_1 = P^{(1)}$  and  $g_2 = P^{(2)}$ , then for any  $F$ —distance generator these formulas coincide with the Formulas (2)–(13) and (65) of classical probabilistic definitions, respectively.*

**Proof.** Let us show the validity of this proposition in the cases of distance, divergence, and similarity relations.

(a) Let us show for a distance that

$$\mathcal{D}^{(F,d)}(P^{(1)}, P^{(2)}) = d(P^{(1)}, P^{(2)}) \quad (71)$$

for any  $F$ —distance generator function.

Consider

$$\mathcal{D}^{(F,d)}(P^{(1)}, P^{(2)}) = F(d(P_1^{(p^{(1)})}, P_1^{(p^{(2)})}), \dots, d(P_m^{(p^{(1)})}, P_m^{(p^{(2)})})). \tag{72}$$

Since the associated probabilities of the probability distribution are unique, that is  $P_i^{(P^{(1)})} = P^{(1)}$ , and  $P_i^{(P^{(2)})} = P^{(2)}$ ,  $i = 1, 2, \dots, m$ , therefore,

$$\mathcal{D}^{(F,d)}(P^{(1)}, P^{(2)}) = F(d(P^{(1)}, P^{(2)}), \dots, d(P^{(1)}, P^{(2)})) = d(P^{(1)}, P^{(2)}),$$

because  $F$  is idempotent.

(b) Let us show for the divergence that

$$\mathcal{D}u_R^{(F,\alpha)}(P^{(1)} \parallel P^{(2)}) = \mathcal{D}u_R^{(\alpha)}(P^{(1)} \parallel P^{(2)}) \tag{73}$$

for any  $F$ —distance generator function. Similarly, to the previous case, we have

$$\begin{aligned} \mathcal{D}u_R^{(F,\alpha)}(P^{(1)} \parallel P^{(2)}) &= F(\mathcal{D}u_R^{(\alpha)}(P_1^{(p^{(1)})} \parallel P_1^{(p^{(2)})}), \dots, \mathcal{D}u_R^{(\alpha)}(P_m^{(p^{(1)})} \parallel P_m^{(p^{(2)})})) = \dots \\ &= F(\mathcal{D}u_R^{(\alpha)}(P^{(1)} \parallel P^{(2)}), \dots, \mathcal{D}u_R^{(\alpha)}(P^{(1)} \parallel P^{(2)})) = \mathcal{D}u_R^{(\alpha)}(P^{(1)} \parallel P^{(2)}) \end{aligned} \tag{74}$$

(c) A similar proposition will hold for Huang’s  $F$ —index  $DSI^{(F)}(\cdot, \cdot)$  (without proof)

$$DSI^{(F)}(P^{(1)}, P^{(2)}) = DSI(P^{(1)}, P^{(2)}) \tag{75}$$

□

#### 4. The Use of Generalized Distance, Divergence, and Similarity Parameters in Fuzzy Measure Identification Problems for $\mathcal{G}^{(2)}(X), \mathcal{G}_{Ch}^{(2)}(X), \mathcal{G}_S(X), \mathcal{G}_{Pos}(X)$ and $\mathcal{G}_{(F,m)}$ Fuzzy Measures Classes

Let us briefly review fuzzy measure identification problems and results that are widely used in interactive MADM models.

In [41], the practical applicability of two probability representations of a finite fuzzy measure—the Campos-Bolanos representation (CBR), equivalent to the APC of a fuzzy measure and the Murofushi-Sugeno representation (MSR)—is examined within the context of multi-criteria decision-making (MCDM) models. This work constructs a new MSR-type representation-interpreter specifically for a particular class of finite fuzzy measures. In [41], a universal interpreter for a capacity (fuzzy measure) in probability MSR, under the Choquet integral framework and second-order dual capacities, is explored. The next research focus is on analyzing fuzzy measure non-additivity indexes, which are relevant for interactive MCDM models where attribute interactions are observed. The non-additivity index effectively measures the degree of attributes interaction. In [42], this index is employed to evaluate the range of advantages concerning decision-maker alternatives. [43] introduces the use of the non-additivity index to replace the Shapley concurrent interaction index and develop an undated MADM decision scheme. A method for calculating the non-additivity index and a decision support algorithm to establish dominance relationships for optimal alternatives ranking are also presented. Key properties of the non-additivity index are considered in [44], along with a capacity identification algorithm based on this index. The algorithm uses linear constraints to reflect decision-maker advantages over alternatives and formulates a linear programming problem to determine the optimal capacity. A capacity identification simulation algorithm based on the non-additivity index is developed in [45]. Another research direction involves the additivity defectiveness of capacities, as discussed in [46]. This paper introduces the concept of capacity defectiveness, representing the degree of capacity non-additivity, and calculates or approximates the defectiveness coefficient for certain capacity classes. An optimal approximation approach for fuzzy integrals is also developed for replacing fuzzy measures with classical measures. The identification of

fuzzy measures, including interaction indexes and importance values (Shapley values), is further examined in [47], where various fuzzy measure representations, such as Möbius transformations and  $k$ -order additive measures, are considered. Ref. [48] shows that every discrete fuzzy measure can be represented as a  $k$ -order additive fuzzy measure, and presents alternative representation methods using interaction indexes and Shapley values. A learning algorithm for identifying  $k$ -maxitive measures based on heuristic least mean squares is given in [49]. Ref. [50] explores the structure and properties of a specific fuzzy measure type applicable to interactive MADM models, utilizing interaction coefficients, Möbius representation and dual fuzzy measures. Ref. [49] introduces the generalized interaction index, or  $g$ -index, which requires significant computational resources, and presents algorithms to calculate the  $g$ -index for  $k$ -maxitive measures. Ref. [51] provides a new visualization scheme for understanding fuzzy measures, while Ref. [52] examines a joint Choquet integral-fuzzy measures operator that uses attribute interactions. Ref. [53] utilizes a hesitant fuzzy linguistic term set to describe attribute interactivity in fuzzy measure identification. Finally, Ref. [47] reviews current approaches to fuzzy measure identification and their advantages and limitations, with Ref. [54] discussing fuzzy measure representations for learning Choquet and Sugeno integrals.

In this section, we present a completely new approach to fuzzy measure identification problems. Conditional optimization problems will be constructed, where the similarity, distance, and divergence parameters defined in the previous sections for two fuzzy measures will be used as objective functions, and the requirements and data of the interactive MADM will be considered in the constraints. The identification fuzzy measure APC will be considered as optimization unknown variables. Specific identification will be performed for specific fuzzy measure classes.

As a first example, let us consider the class  $\mathcal{G}^{(2)}(X)$  of two-additive fuzzy measures. Suppose we are given an interactive MADM model [55,56] with possible alternatives  $D = (d_1, d_2, \dots, d_m)$  and attributes  $X = \{x_1, x_2, \dots, x_n\}$ . Suppose that in the interactive MADM model, attributes importance values  $\{I_1, I_2, \dots, I_n\}$  and pairwise interactions are known in the form of a symmetric matrix  $\{I_{ij}\}_{i=i, \dots, n; j=1, \dots, n', i \neq j}$  [57–62]. As is known in [26], if  $\hat{g} \in \mathcal{G}^{(2)}(X)$ , then its associated probabilities are calculated by the following formula:  $\forall \sigma \in S_n$

$$\hat{P}_\sigma(x_{\sigma(i)}) \equiv P_\sigma^{(\hat{g})}(x_{\sigma(i)}) = I_{\sigma(i)} + \frac{1}{2} \sum_{j=1}^{i-1} I_{\sigma(i)\sigma(j)} - \frac{1}{2} \sum_{j=i+1}^n I_{\sigma(i)\sigma(j)}, \quad i = 1, \dots, n. \quad (76)$$

In (76), if  $i = 1$ , then the second addend is equal to 0, and if  $i = n$ , then the third addend is equal to 0. Assume that in the decision-making matrix there are alternatives with ratings  $\{\xi_{ij}\}$  from  $[0, 1]$  (see Table 1):

**Table 1.** MADM decision-making matrix.

d	x					
	$x_1x_2$	...	$x_j$	...	$x_n$	
$d_1$						
$d_2$						
...						
$d_i$	$\xi_{i1}\xi_{i2}$	...	$\xi_{ij}$	...	$\xi_{in}$	
...						
$d_m$						

Suppose that the index of uncertainty of expert evaluations on attributes  $X = \{x_1, x_2, \dots, x_n\}$  is represented by  $g \in \mathcal{G}(X)$  fuzzy measure. If there is a dependence between the attributes in the form of interactions, then it is completely acceptable to use the Choquet finite integral as an aggregation tool [20]. Then, to obtain a scalar evaluation of the ranking of each alternative  $d_i$  and to aggregate its ratings  $(\xi_{i1}, \xi_{i2}, \dots, \xi_{ij}, \dots, \xi_{in})$  to make a final decision, we will use the value

of the Choquet integral with respect to the fuzzy measure  $g$  [20]. We highlight the following publications of the authors of this study in the direction of using associated probabilities based on extensions of the Choquet and Sugeno integral operators [63,64]. We also know that this population parameter is known in fuzzy statistics as monotone expectation ( $ME(d_i)$ ) [23], i.e.,

$$ME(d_i) = \int_0^{+\infty} g(\{x_j \in X, \zeta_{ij} \geq \alpha\})d\alpha, \quad i = 1, 2, \dots, m, \tag{77}$$

or

$$ME(d_i) = \sum_{j=1}^n P_{\tau}^{(g)}(x_{\tau(j)}) \cdot \zeta_{i\tau(j)}, \quad i = 1, 2, \dots, m, \tag{78}$$

where  $\tau = (\tau(1), \dots, \tau(m)) \in S_n$  is a permutation for which  $1 \geq \zeta_{i\tau(1)} \geq \zeta_{i\tau(2)} \geq \dots \geq \zeta_{i\tau(n)} \geq 0$ .

Now let us consider the concepts of the best approximation of a particular fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  from a given class of fuzzy measures. The idea is that this approximation is achieved for such fuzzy measure  $\hat{g}$  from the given fuzzy measures subclass that the distance or divergence to a particular fuzzy measure  $\bar{g}$  is the smallest or this approximation is achieved for such fuzzy measure  $\hat{g}$  from the subclass of given fuzzy measures that the similarity to the fuzzy measure  $\bar{g}$  is greatest. In doing so, we determine a similar to concrete approximate fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  from the given class of fuzzy measures. Consider Definitions (16)–(19), which are different cases of approximation.

**Definition 16.** The fixed two-additive fuzzy measure  $\hat{g} \in \mathcal{G}^{(2)}(X)$  is called the  $F$ —best approximation of the fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  according to the similarity relation for the  $F$ —distance generator

$$DSI^{(F)}(\bar{g}, \hat{g}) = \sup_{g \in \mathcal{G}(X)} DSI^{(F)}(g, \hat{g}). \tag{79}$$

Therefore, the best, highly similar fuzzy measure  $\bar{g}$  among the two-additive fuzzy measures is the known fuzzy measure  $\hat{g}$ .

Analogous definitions can be made in relation to the distance and divergence between fuzzy measures.

**Definition 17.** A fixed two-additive fuzzy measure  $\hat{g} \in \mathcal{G}^{(2)}(X)$  is called the  $F$ —best approximation of the fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  according to the  $\mathcal{D}^{(F,d)}$ —distance relation for the  $F$ —distance generator

$$\mathcal{D}^{(F,d)}(\bar{g}, \hat{g}) = \inf_{g \in \mathcal{G}(X)} \mathcal{D}^{(F,d)}(g, \hat{g}). \tag{80}$$

Note that the  $\mathcal{D}^{(F,d)}$ —distance in Definition 17 can be replaced by the  $\mathcal{D}_R^{(F)}$ —distance defined above or other distances generalized here.

**Definition 18.** The fixed two-additive fuzzy measure  $\hat{g} \in \mathcal{G}^{(2)}(X)$  is called a  $F$ —best approximation of the fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  according to the  $\mathcal{D}u_R^{(F,\alpha)}$ —divergence for the  $F$ —distance generator

$$\mathcal{D}u_R^{(F,\alpha)}(\bar{g} \parallel \hat{g}) = \inf_{g \in \mathcal{G}(X)} \mathcal{D}u_R^{(F,\alpha)}(g \parallel \hat{g}). \tag{81}$$

Note here again that the  $\mathcal{D}u_R^{(F,\alpha)}$ —divergence in Definition 18 can be replaced by divergences  $\mathcal{D}u_{KL}^{(F)}$ ,  $\mathcal{D}u_J^{(F)}$ ,  $\mathcal{D}u_{JS}^{(F)}$  defined above or other generalized ones mentioned here.

Note that the area of approximation  $\mathcal{G}^{(2)}(X)$  in Definitions 16, 17, and 18 can be replaced by other practically important subclasses of fuzzy measures and presented here:  $\mathcal{G}_S(X)$ —Sugeno  $\lambda$ —additive fuzzy measures,  $\mathcal{G}_{Ch}^{(2)}(X)$ —Choquet second-order capacity,  $\mathcal{G}_{Pos}(X)$ —possibility measures,  $\mathcal{G}_{(F,m)}(X)$ —fuzzy measures associated with the body of evidence and other classes. Obviously, such subclasses of fuzzy measures are considered, for which their associated probability classes are derived, whose representations include only the basic data of these classes. For example, formulas: For the possibility measures (31); for the Sugeno  $\lambda$ —additive measure (17); for the measure associated with the data body (37). Therefore, let us introduce analogous definitions for these fuzzy measures’ classes.

**Definition 19.** The fixed  $\hat{g}_\lambda \in \mathcal{G}_S(X)$ —Sugeno  $\lambda$ —additive fuzzy measure is called the  $F$ —best approximation of the fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  according to the  $\mathcal{D}^{(F,d)}$ —distance relation (according to the  $\mathcal{D}^{(F)}$ —divergence relation), if for the  $F$ —distance generator

$$\mathcal{D}^{(F,d)}(\bar{g}, \hat{g}_\lambda) = \inf_{g \in \mathcal{G}(X)} \mathcal{D}^{(F,d)}(g, \hat{g}_\lambda) \left( \mathcal{D}^{(F)}(\bar{g}, \hat{g}_\lambda) = \inf_{g \in \mathcal{G}(X)} \mathcal{D}^{(F)}(g, \hat{g}) \right) . \tag{82}$$

Let us now introduce a completely new type of fuzzy measure identification problem in the MADM environment based on the definitions given here. In general, the idea is as follows: a scalar problem of conditional optimization will be constructed, the objective function of which will be the concept of one of the approximations. The problem constraints are constructed from the particular MADM problem data and constraints, and the optimization unknown variables will be the associated probabilities from the APC which should be identified; specifically Definitions 16–18 in the MADM environment allow us to construct the model uncertainty index—fuzzy measure  $\bar{g}$  with the best approximation from the given fuzzy measures class, while taking into account other MADM data as constraints. For example, from a practical point of view it is possible to create an approximate interval evaluation of their monotone expectation for some alternatives (with confidence intervals and others). Consider a particular problem with respect to the DSI index when considering a class  $\mathcal{G}^{(2)}(X)$ . To solve this problem, let us make a conditional optimization problem, in which the associated probability class  $\{P_\sigma(\cdot)\}_{\sigma \in S_n}$  of the fuzzy measure  $\bar{g}$  will be unknown:

$$\left\{ \begin{array}{l} DSI^{(F)}(g, \hat{g}) = F(DSI(P_\sigma, \hat{P}_\sigma), \sigma \in S_n) \Rightarrow \max \\ \text{S.C. } \sum_{i=1}^n P_\sigma(x_i), \sigma \in S_n, \\ 0 \leq P_\sigma(x_i) \leq 1, \sigma \in S_n, i = 1, 2, \dots, n, \\ M_{i_k}^- \leq ME(d_{i_k}) = \sum_{j=1}^n P_{\tau_k}(x_{\tau_k(j)}) \cdot \xi_{i_k \tau_k(j)} \leq M_{i_k}^+, k = 1, 2, \dots, s, \end{array} \right. \tag{83}$$

where  $ME(d_{i_k})$  are the monotone expected values for the alternatives  $d_{i_1}, d_{i_2}, \dots, d_{i_s}$  evaluated at intervals by the experts. In the formulas of (83) for each  $i_k \exists \tau_k \in S_n$ , that  $\xi_{i_k \tau_k(1)} \geq \xi_{i_k \tau_k(2)} \geq \dots \geq \xi_{i_k \tau_k(n)}$ .

The conditional optimization problem’s (83) solution fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  is the closest, similar fuzzy measure to the second-order additive capacity  $\hat{g}$ , which takes into account the constraints associated with probabilities in the form of attribute interaction indexes and MADM data on expert evaluations of expected ratings of alternatives. Consider the numerical example of (83):

**Example 1.** Let  $F = F_1$ ,  $n = 3$  and  $m = 4$ . Suppose that the decision-making matrix is as follows (Table 2):

**Table 2.** Decision matrix.

$d_i \backslash x_j$	$x_1$	$x_2$	$x_3$	$ME(d_i)$
$d_1$	0.3	0.7	0.5	-
$d_2$	0.6	0.4	0.8	$0.45 \leq ME(d_2) \leq 0.65$
$d_3$	0.5	0.6	0.7	-
$d_4$	0.8	0.7	0.4	$0.35 \leq ME(d_4) \leq 0.55$

Assume that the importance values  $\{I_j\}$  of the attributes and the indexes of their pairwise interaction  $\{I_{ij}\}$  are as follows (Table 3):

**Table 3.** Magnitudes of attribute values  $\{I_j\}$  and pairwise interaction indexes  $\{I_{ij}\}$ .

$I_{ij}$	$x_1$	$x_2$	$x_3$	$I_j$
$x_1$	-	0.15	0.25	0.3
$x_2$	0.15	-	0.10	0.4
$x_3$	0.25	0.10	-	0.3

Using Table 3 and Formulas (76), calculate the associated probabilities class  $\{\hat{P}_\sigma(\cdot)\}_{\sigma \in S_3}$  of the fuzzy measure  $\hat{g} \in \mathcal{G}^{(2)}(X)$  (see Table 4).

**Table 4.** The class of associated probabilities  $\{\hat{P}_\sigma(\cdot)\}_{\sigma \in S_3}$ .

$\sigma \backslash x_j$	$\hat{P}_\sigma(\cdot)$	$x_1$	$x_2$	$x_3$
$\sigma_1 = (1, 2, 3)$	$\hat{P}_{\sigma_1}(\cdot)$	0.100	0.425	0.475
$\sigma_2 = (1, 3, 2)$	$\hat{P}_{\sigma_2}(\cdot)$	0.100	0.525	0.375
$\sigma_3 = (2, 1, 3)$	$\hat{P}_{\sigma_3}(\cdot)$	0.250	0.275	0.475
$\sigma_4 = (2, 3, 1)$	$\hat{P}_{\sigma_4}(\cdot)$	0.500	0.275	0.225
$\sigma_5 = (3, 1, 2)$	$\hat{P}_{\sigma_5}(\cdot)$	0.350	0.525	0.125
$\sigma_6 = (3, 2, 1)$	$\hat{P}_{\sigma_6}(\cdot)$	0.500	0.375	0.125

Then (83) will take a specific form with respect to the unknown associated probabilities values  $\{P_\sigma(\cdot)\}$ :

$$\left\{ \begin{array}{l}
 DSI^{(F_1)}(g, \hat{g}) = \frac{2}{3!} \cdot \frac{\sum_{j=1}^6 \sum_{i=1}^3 [P_{\sigma_j}(x_i) \cdot \hat{P}_{\sigma_j}(x_i)]}{\sum_{j=1}^6 \sum_{i=1}^3 [(P_{\sigma_j}(x_i))^2 + (\hat{P}_{\sigma_j}(x_i))^2]} \Rightarrow \max \\
 \text{S.C.} \quad 0 \leq P_{\sigma_j}(x_i) \leq 1, \quad j = 1, 2, \dots, 6, \quad i = 1, 2, 3, \\
 \sum_{i=1}^3 P_{\sigma_j}(x_i) = 1, \quad j = 1, 2, \dots, 6, \\
 0.45 \leq \sum_{i=1}^3 P_{\sigma_5}(x_{\sigma_5(i)}) \cdot \zeta_{2\sigma_5(i)} = ME(d_2) \leq 0.65, \quad \sigma_5 = (3, 1, 2), \\
 0.35 \leq \sum_{i=1}^3 P_{\sigma_1}(x_{\sigma_1(i)}) \cdot \zeta_{4\sigma_1(i)} = ME(d_4) \leq 0.55, \quad \sigma_1 = (1, 2, 3),
 \end{array} \right. \tag{84}$$

where

$$\begin{aligned}
 (\zeta_{2\sigma_5(1)}, \zeta_{2\sigma_5(2)}, \zeta_{2\sigma_5(3)}) &= (0.8, 0.6, 0.4), \\
 (\zeta_{4\sigma_1(1)}, \zeta_{4\sigma_1(2)}, \zeta_{4\sigma_1(3)}) &= (0.8, 0.7, 0.4).
 \end{aligned}$$

The numerical solution of the conditional optimization problem (84) as an associated probabilities class of the fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  is presented in Table 5.

**Table 5.** Associated probabilities class of the fuzzy measure  $\bar{g}$  (for the problem (84)).

$\sigma$	$x_i$	$P_{\sigma}(\cdot)$	$x_1$	$x_2$	$x_3$
$\sigma_1 = (1, 2, 3)$		$P_{\sigma_1}(\cdot)$	0.0855	0.4398	0.4747
$\sigma_2 = (1, 3, 2)$		$P_{\sigma_2}(\cdot)$	0.1103	0.5322	0.3575
$\sigma_3 = (2, 1, 3)$		$P_{\sigma_3}(\cdot)$	0.2500	0.2750	0.4750
$\sigma_4 = (2, 3, 1)$		$P_{\sigma_4}(\cdot)$	0.5000	0.2750	0.2250
$\sigma_5 = (3, 1, 2)$		$P_{\sigma_5}(\cdot)$	0.3500	0.5250	0.1250
$\sigma_6 = (3, 2, 1)$		$P_{\sigma_6}(\cdot)$	0.5000	0.3750	0.1250

The maximum similarity index is equal to  $DSI(\bar{g}, \hat{g}) = 0.89997$ .

Of course, it was possible to use any distance or divergence Formulas (47)–(52) or (53)–(64) as the objective function in the problem (83)–(84). Additionally, we could use the distances  $\mathcal{D}_2, \mathcal{D}_B^{(F)}(\mathcal{D}_B^{(F_q)}, \mathcal{D}_B^{(F_{max})})$  as a minimization criterion. Also, we could use the divergences  $\mathcal{D}u_R^{(F,\alpha)}, \mathcal{D}u_{KL}^{(F)}(\mathcal{D}u_{KL}^{(F_q)}, \mathcal{D}u_{KL}^{(F_{max})}), \mathcal{D}u_J^{(F)}(\mathcal{D}u_J^{(F_{max})}, \mathcal{D}u_J^{(F_q)}),$  and  $\mathcal{D}u^{(F,\alpha)}, \mathcal{D}u_{JS}^{(F)}(\mathcal{D}u_{JS}^{(F_q)}, \mathcal{D}u_{JS}^{(F_{max})})$  as a minimization criterion. Consider another example:

**Example 2.** We consider the conditional optimization problem (83) and (84), but we choose the Jeffrey’s  $F_1$ —divergence as the objective function, and the constraints will be the same. MADM data are not changed, that is, the class of associated probabilities  $\{\hat{P}_{\sigma}(\cdot)\}_{\sigma \in S_3}$  is the same. We get the following conditional optimization problem:

$$\left\{ \begin{array}{l} \mathcal{D}u_J^{(F_1)}(g, \hat{g}) = \frac{1}{3!} \sum_{j=1}^6 \left( \sum_{i=1}^3 [P_{\sigma_j}(x_i) - \hat{P}_{\sigma_j}(x_i)] \cdot \log \left[ \frac{P_{\sigma_j}(x_i)}{\hat{P}_{\sigma_j}(x_i)} \right] \right) \Rightarrow \min \\ \text{S.C.} \quad \sum_{i=1}^3 P_{\sigma_j}(x_i) = 1, \quad j = 1, 2, \dots, 6 \\ 0 \leq P_{\sigma_j}(x_i) \leq 1, \quad j = 1, 2, \dots, 6; \quad i = 1, 2, 3 \\ 0.45 \leq \sum_{i=1}^3 P_{\sigma_5}(x_{\sigma_5(i)}) \cdot \xi_{2\sigma_5(i)} \leq 0.65 \\ 0.35 \leq \sum_{i=1}^3 P_{\sigma_1}(x_{\sigma_1(i)}) \cdot \xi_{4\sigma_1(i)} \leq 0.55 \end{array} \right. \quad (85)$$

Table 6 presents the results of the numerical solution of the conditional optimization problem (85) in the form of the class of associated probabilities of the fuzzy measure  $\bar{g} \in \mathcal{G}(X)$ .

The minimum divergence value is equal to  $\mathcal{D}u_J^{(F_1)}(\bar{g}, \hat{g}) = 0.00021$ .

Of course, for the identification of the associated probabilities class of a fuzzy measure, one could also consider multi-objective conditional optimization problems by considering several objective functions. This should be derived from the issues of the synthesis of a specific research problem.

**Table 6.** The class of associated probabilities of fuzzy measure  $\bar{g}$  for problem (85).

$\sigma$	$x_i$	$P_\sigma(\cdot)$	$x_1$	$x_2$	$x_3$
$\sigma_1 = (1, 2, 3)$		$P_{\sigma_1}(\cdot)$	0.1052	0.4203	0.4745
$\sigma_2 = (1, 3, 2)$		$P_{\sigma_2}(\cdot)$	0.1002	0.5249	0.3749
$\sigma_3 = (2, 1, 3)$		$P_{\sigma_3}(\cdot)$	0.2500	0.2749	0.4751
$\sigma_4 = (2, 3, 1)$		$P_{\sigma_4}(\cdot)$	0.5000	0.2749	0.2251
$\sigma_5 = (3, 1, 2)$		$P_{\sigma_5}(\cdot)$	0.3500	0.5250	0.1250
$\sigma_6 = (3, 2, 1)$		$P_{\sigma_6}(\cdot)$	0.5000	0.3749	0.1251

Consider one last example:

**Example 3.** Consider another example. Let us choose the domain of approximation of a fuzzy measure  $\bar{g}$  as all classes of fuzzy measures  $\mathcal{G}_{\langle \mathcal{F}, m \rangle}(X)$  associated with the body of data  $\langle \mathcal{F}, m \rangle$  defined on  $2^X$ , where for each fuzzy measure  $\bar{g}$  its associated probabilities are calculated as follows (Formula (38)),  $\forall \sigma \in S_n$ :

$$P_\sigma(x_{\sigma(i)}) = \sum_{A_j \in \mathcal{F}: A_j \cap \{x_{\sigma(i)}\} \neq \emptyset} m(A_j) \cdot w_j(|A_j \cap \{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}|), \quad i = 1, \dots, n. \quad (86)$$

As the index of the difference between the approximation fuzzy measure  $\bar{g}$  and the fuzzy measure  $\hat{g}$  for the objective function of the corresponding conditional optimization problem, let us choose the Jeffrey’s symmetric  $F_2$ —divergence  $Du_j^{(F_2)}(\cdot, \cdot)$ :

$$Du_j^{(F_2)}(g, \hat{g}) = \sqrt{\frac{1}{n!} \left( \sum_{\sigma \in S_n} \left( \sum_{i=1}^n [P_\sigma(x_i) - \hat{P}_\sigma(x_i)] \cdot [\log_2 P_\sigma(x_i) - \log_2 \hat{P}_\sigma(x_i)] \right)^2 \right)}. \quad (87)$$

To construct a numerical example, let us take the same matrix from Table 2 as the decision-making matrix of the MADM model. Consider the Choquet integral with respect to the fuzzy measure  $\bar{g}$  as an aggregation operator. Consider the following body of evidence  $\langle \mathcal{F}, m \rangle$ :  $\mathcal{F} = \{A_1, A_2\}$ ,  $A_1 = \{x_1, x_2\}$ ,  $A_2 = \{x_2, x_3\}$ ,  $m(A_1) = \frac{2}{3}$ ,  $m(A_2) = \frac{1}{3}$  and the vector of allocation weights  $W = \{w_1, w_2\}$ ,  $w_1 = \{w_1(1), w_1(2)\} = \left\{ \frac{3}{4}, \frac{1}{4} \right\}$ ;  $w_2 = \{w_2(1), w_2(2)\} = \left\{ \frac{1}{5}, \frac{4}{5} \right\}$ . Table 7 presents the fuzzy measure  $\hat{g}$  associated with this body of evidence and weights, and more precisely, its associated probability class  $\{\hat{P}_\sigma\}_{\sigma \in S_3}$ , for this we used Formula (38).

**Table 7.** Class of associated probabilities  $\{\hat{P}_\sigma\}_{\sigma \in S_3}$  of the fuzzy measure  $\hat{g}$  associated with the body of evidence  $\langle \mathcal{F}, m \rangle$  and the given allocation weights vector  $W = \{w_1, w_2\}$ .

$\sigma$	$x_i$	$\hat{P}_\sigma(\cdot)$	$x_1$	$x_2$	$x_3$
$\sigma_1 = (1, 2, 3)$		$\hat{P}_{\sigma_1}(\cdot)$	0.500	0.233	0.267
$\sigma_2 = (1, 3, 2)$		$\hat{P}_{\sigma_2}(\cdot)$	0.500	0.433	0.067
$\sigma_3 = (2, 1, 3)$		$\hat{P}_{\sigma_3}(\cdot)$	0.167	0.567	0.267
$\sigma_4 = (2, 3, 1)$		$\hat{P}_{\sigma_4}(\cdot)$	0.167	0.567	0.267
$\sigma_5 = (3, 1, 2)$		$\hat{P}_{\sigma_5}(\cdot)$	0.500	0.433	0.067
$\sigma_6 = (3, 2, 1)$		$\hat{P}_{\sigma_6}(\cdot)$	0.167	0.767	0.067

To identify the fuzzy measure  $\bar{g}$  as the MADM uncertainty index, consider the conditional optimization problem with the constraints of the previous examples and the divergence  $\mathcal{D}_{\nu_j^{(F_2)}}(\cdot, \cdot)$  minimization criterion:

$$\left\{ \begin{aligned} \mathcal{D}_{\nu_j^{(F_2)}}(g, \hat{g}) &= \sqrt{\frac{1}{3!} \left( \sum_{j=1}^6 \sum_{i=1}^3 [P_{\sigma_j}(x_i) - \hat{P}_{\sigma_j}(x_i)] \cdot [\log_2 P_{\sigma_j}(x_i) - \log_2 \hat{P}_{\sigma_j}(x_i)] \right)^2} \Rightarrow \min \\ \text{S.C.} \quad &\sum_{i=1}^3 P_{\sigma_j}(x_i) = 1, \quad j = 1, 2, \dots, 6 \\ &0 \leq P_{\sigma_j}(x_i) \leq 1, \quad j = 1, 2, \dots, 6; \quad i = 1, 2, 3 \\ &0.45 \leq \sum_{i=1}^3 P_{\sigma_5}(x_{\sigma_5(i)}) \cdot \xi_{2\sigma_5(i)} \leq 0.65 \\ &0.35 \leq \sum_{i=1}^3 P_{\sigma_1}(x_{\sigma_1(i)}) \cdot \xi_{4\sigma_2(i)} \leq 0.55 \end{aligned} \right. \quad (88)$$

The associated probability class of the fuzzy measure  $\bar{g} \in \mathcal{G}(X)$  for the solution of the problem (88) is given in Table 8:

**Table 8.** The associated probability class of the approximation fuzzy measure  $\bar{g}$ .

$\sigma$	$x_i$	$P_{\sigma}(\cdot)$	$x_1$	$x_2$	$x_3$
$\sigma_1 = (1, 2, 3)$		$P_{\sigma_1}(\cdot)$	0.5218	0.2411	0.2371
$\sigma_2 = (1, 3, 2)$		$P_{\sigma_2}(\cdot)$	0.4953	0.4332	0.0715
$\sigma_3 = (2, 1, 3)$		$P_{\sigma_3}(\cdot)$	0.1839	0.5478	0.2683
$\sigma_4 = (2, 3, 1)$		$P_{\sigma_4}(\cdot)$	0.1678	0.5636	0.2686
$\sigma_5 = (3, 1, 2)$		$P_{\sigma_5}(\cdot)$	0.4991	0.4457	0.0552
$\sigma_6 = (3, 2, 1)$		$P_{\sigma_6}(\cdot)$	0.1714	0.7632	0.0654

The minimum value of the objective function  $\mathcal{D}_{\nu_j^{(F_2)}}(g, \hat{g})$  is:  $\mathcal{D}_{\nu_j^{(F_2)}}(\bar{g}, \hat{g}) = 0.00506$ .

For all three examples, at the last stage, we complete the MADM problem presented in Table 2 and rank the alternatives  $\langle d_1, d_2, d_3, d_4 \rangle$  using the identified fuzzy measure  $\bar{g}$ . Aggregation will be accomplished again with the Choquet integral

$$ME(d_i) = \sum_{j=1}^3 P_{\tau^{(i)}}(x_{\tau^{(i)}(j)}) \cdot \xi_{i\tau^{(i)}(j)}, \quad i = 1, 2, 3, 4,$$

where for each alternative  $d_i$  there is a permutation  $\tau^{(i)} = (\tau^{(i)}(1), \tau^{(i)}(2), \tau^{(i)}(3)) \in S_3$  such that

$$\xi_{i\tau^{(i)}(1)} \geq \xi_{i\tau^{(i)}(2)} \geq \xi_{i\tau^{(i)}(3)}.$$

Table 9 summarizes the aggregation values for all three examples, and Table 10 presents the ranking of the alternatives.

**Table 9.** Values of aggregations on alternatives  $\langle d_1, d_2, d_3, d_4 \rangle$  for all three examples.

	Aggregation Values for Example 1	Aggregation Values for Example 2	Aggregation Values for Example 3
$d_1$	0.461	0.454	0.578
$d_2$	0.529	0.529	0.520
$d_3$	0.569	0.563	0.590
$d_4$	0.577	0.570	0.680

**Table 10.** Rankings of alternatives  $\langle d_1, d_2, d_3, d_4 \rangle$ .

	Alternative Rankings
Example 1	$d_4 \succ d_3 \succ d_2 \succ d_1$
Example 2	$d_4 \succ d_3 \succ d_2 \succ d_1$
Example 3	$d_4 \succ d_3 \succ d_1 \succ d_2$

It becomes clear that the alternative  $d_4$  is the best for all problems.

Comparative Analysis: We note again that the examples given here are for illustrative purposes only and their purpose is to make it easy for the reader to construct the analogous optimization model they need. However, a little comparative analysis can be conducted. Examples 1 and 2 are practically indistinguishable examples in that the MADM environment is the same, only the objective functions differ. e.g., In 1, the objective function is the similarity parameter  $DSI^{(F_1)}(\cdot, \cdot)$  and its maximization is considered. However, the maximum similarity parameter is quite high— $DSI^{(F_1)} = 0.89997$ . And, e.g., in 2, the objective function is the divergence parameter  $\mathcal{D}v_j^{(F_2)}(\cdot, \cdot)$  and its minimization is considered. Here too, the minimum divergence parameter is quite high— $\mathcal{D}v_j^{(F_2)} = 0.00021$ . Both examples consider interactive MADM in the approximation of pairwise interaction of environmental attributes, where the two-additive fuzzy measure is considered as the uncertainty index (Table 4). In both examples, the approximation class is the same, and the similarity between the identified fuzzy measures is high (Tables 5 and 6). Therefore, the ranking of the alternatives should be almost identical (in our case it is completely identical (Table 10).

As for the third example, here the MADM environment is partially different from the case of the previous two examples. Here, the uncertainty index is considered to be the body of evidence and the class of fuzzy measures associated with it (86). A specific allocation weight vector and a specific body of evidence are considered. Probabilistic representations of the associated fuzzy measure are given in Table 7. There is agreement in all three examples, where all conditional optimization problems' constraints are the same. In the third example, the objective function is the same as the divergence parameter  $\mathcal{D}v_j^{(F_2)}(\cdot, \cdot)$  of the second example, and its minimization is considered. Here too, the minimum divergence parameter is quite high— $\mathcal{D}v_j^{(F_2)} = 0.00506$ . As a result, the ranking of the alternatives in the third example partially coincided with the rankings of the alternatives in the first and second examples.

We tested different divergence or similarity parameters in the objective function role for the same MADM environment. We also changed the environment of MADM. As a result, the rankings of the alternatives for the fixed MADM environment are almost identical when we take the Choquet finite integral as the aggregation operator. Which indicates the sensitivity of the identification problem constructed in the article.

### 5. Conclusions

The presented paper discusses the binary relations of distance, divergence and similarity defined on the space of finite fuzzy measures, which in a certain sense represent generalizations of similar binary relations defined on the space of finite probability distributions. The correctness of the generalizations is proved. The generalizations are based on the class of fuzzy measure—associated probabilities. More precisely, the distance, divergence, and similarity parameters between two fuzzy measures are determined by the distance, divergence, and similarity parameters between their associated probability classes. For the definition of the mentioned parameters between the associated probability classes, the concept of distance generator is introduced, which scales the values of the associated probability class in the scalar value of the corresponding parameter.

The concepts of distance, divergence and similarity between fuzzy measures are used in the fuzzy measure identification problem for a certain multi-attribute decision-making

(MADM) environment. For this, a conditional optimization problem with a single objective function that actually represents a parameter of distance, divergence, or similarity, is formulated. The constraints of the optimization problem represent constraints on the MADM data, as well as constraints on the associated probabilities of identification fuzzy measure. Given the extrema of the objective function, the identification fuzzy measure is the best approximation for the class of fuzzy measures allowed in the MADM environment.

The classes of the second-order Choquet capacities, two-additive fuzzy measures, Sugeno  $\lambda$ —additive fuzzy measures, possibility measures, and fuzzy measures associated with the body of evidence are considered. Numerical examples are discussed and a comparative analysis of the obtained results is presented. In the conditional optimization problem corresponding to a simple example of three parallel MADMs, one and the same constraints are considered. The only difference is in the allowable classes of the search fuzzy measure, which are associated with the nature of the specific problem. There is also a difference in the selection of the objective function, which derives from the choice of preferences of the decision-maker.

The results are obtained at the level of ranking of MADM alternatives. For ranking, data aggregations of alternatives are created using the fuzzy measure identified in it from the Choquet aggregation operator. The results show some differences, which are due to the selection of the search class of the fuzzy measure approximation and the preferences of the decision-maker in the selection of the objective function.

We have considered the same examples in the case of the Sugeno finite integral. The ranking of the alternatives was found to be in great agreement with the case of the Choquet integral. It would also be interesting to consider other aggregation integral operators that use fuzzy measures in their calculations.

Future studies will consider other classes of fuzzy measure approximations related to specific important practical problems. Other important new divergence and distance generalizations for fuzzy measures that are not within the scope of this article will also be considered. Changing the aggregation operator was sensitive.

In future studies, we will consider the use of metrics, similarity and divergence relations defined in the fuzzy measures space in aggregation relations with non-additive and interacting parameters or attributes, such as users' similarity relation in collaborative filtering systems, phase space metrics of non-additive components of machine learning, measures of clustering and classification of complex objects, etc. In our future studies, it would also be natural to describe the relationships between the divergences mentioned here and the entropy measures.

**Author Contributions:** The authors contributed equally in this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Shota Rustaveli National Scientific Foundation of Georgia (SRNSF), grant number [FR-22-969].

**Data Availability Statement:** The paper is original and, therefore, no data were used.

**Acknowledgments:** The authors are grateful to the anonymous reviewers for their valuable comments and suggestions in improving the quality of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mahalanobis, P.C. On the generalized distance in statistics. *Sankhya Indian J. Stat. Ser. A* **1936**, *80*, 1–7.
2. Bhattacharyya, A. On discrimination and divergence. In Proceedings of the India Science Congress, Asiatic Society of Bengal, Calcutta, India, 1942. Available online: [https://books.google.com.sg/books/about/Proceedings\\_of\\_the\\_Indian\\_Science\\_Congre.html?id=9mYbAAAAMAAJ&redir\\_esc=y](https://books.google.com.sg/books/about/Proceedings_of_the_Indian_Science_Congre.html?id=9mYbAAAAMAAJ&redir_esc=y) (accessed on 15 August 2024).
3. Nikulin, M.S. Hellinger distance 1994. In *Encyclopedia of Mathematics*; EMS Press, 2001, Available online: [https://en.wikipedia.org/wiki/Hellinger\\_distance](https://en.wikipedia.org/wiki/Hellinger_distance) (accessed on 15 August 2024).
4. Bhattacharyya, A. On a Measure of Divergence between Two Multinomial Populations. *Sankhyā* **1946**, *7*, 401–406.
5. Wikipedia, F-Divergence. Available online: <https://en.wikipedia.org/wiki/F-divergence> (accessed on 15 August 2024).

6. Rényi, A. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, Los Angeles, CA, USA, 30 June–30 July 1960; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.
7. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
8. Csiszar, I.  $\mathcal{I}$ -Divergence Geometry of Probability Distributions and Minimization Problems. *Ann. Probab.* **1975**, *3*, 146–158. [[CrossRef](#)]
9. Jeffreys, H. *Theory of Probability*, 2nd ed.; Oxford University Press: Oxford, UK, 1948.
10. Schütze, H.; Manning, C.D. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, UK, 1999; p. 304.
11. Sterreicher, F.; Vajda, I. A new class of metric divergences on probability spaces and its statistical applications. *Ann. Inst. Statist. Math.* **2003**, *55*, 639–653. [[CrossRef](#)]
12. Nielsen, F. On a variational definition for the Jensen-Shannon summarization of distances based on the information radius. *Entropy* **2021**, *23*, 464. [[CrossRef](#)]
13. Liang, Z.; Jianhui, L. The Similarity for Nominal Variables Based on F-Divergence. *Int. J. Database Theory Appl.* **2016**, *9*, 191–202. [[CrossRef](#)]
14. Deng, J.; Wang, Y.; Guo, J.; Deng, Y.; Park, Y. A similarity measure based on Kullback–Leibler divergence for collaborative filtering in sparse data. *J. Inf. Sci.* **2018**, *45*, 656–675. [[CrossRef](#)]
15. Pastore, M.; Calcagni, A. Measuring Distribution Similarities Between Samples: A Distribution-Free Overlapping Index. *Front. Psychol.* **2019**, *10*, 1089. [[CrossRef](#)]
16. Cai, Y.; Lim, L.-H. Distances Between Probability Distributions of Different Dimensions. *IEEE Trans. Inf. Theory* **2022**, *68*, 4020–4031. [[CrossRef](#)]
17. Le, H.; Sang, V.N.T.; Thuy, L.N.L.; Bao, P.T. The fuzzy Kullback–Leibler divergence for estimating parameters of the probability distribution in fuzzy data: An application to classifying Vietnamese Herb Leaves. *Sci. Rep.* **2023**, *13*, 14537. [[CrossRef](#)]
18. Fei, L.; Deng, Y. A new divergence measure for basic probability assignment and its applications in extremely uncertain environments. *Int. J. Intell. Syst.* **2018**, *34*, 584–600. [[CrossRef](#)]
19. Rachev, S.T.; Klebanov, L.B.; Stoyanov, S.V.; Fabozzi, F. *The Methods of Distances in the Theory of Probability and Statistics*; Springer: New York, NY, USA, 2013; p. 619.
20. Choquet, G. Theory of Capacities. *Ann. L’Institut Fourier* **1953**, *5*, 131–295.
21. Sugeno, M. Theory of Fuzzy Integrals and Its Applications. Ph.D. Thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.
22. Denneberg, D. *Non-Additive Measure and Integral*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1994. [[CrossRef](#)]
23. Campos, L.M.; Bolanos, M.N. Representation of fuzzy measures through probabilities. *Fuzzy Set Syst.* **1989**, *31*, 23–36.
24. Sirbiladze, G. New Fuzzy Aggregation Operators Based on the Finite Choquet Integral—Application in the MADM Problem. *Int. J. Inf. Technol. Decis. Mak.* **2016**, *15*, 517–551. [[CrossRef](#)]
25. Sirbiladze, G.; Khutsishvili, I.; Midodashvili, B. Associated immediate probability intuitionistic fuzzy aggregations in MCDM. *Comput. Ind. Eng.* **2018**, *123*, 1–8. [[CrossRef](#)]
26. Sirbiladze, G. Associated Probabilities in Interactive MADM under Discrimination q-Rung Picture Linguistic Environment. *Mathematics* **2021**, *9*, 2337. [[CrossRef](#)]
27. Sirbiladze, G.; Garg, H.; Khutsishvili, I.; Ghvaberidze, B.; Midodashvili, B. Associated probabilities aggregations in multistage investment decision-making. *Kybernetes* **2021**, *52*, 1370–1399. [[CrossRef](#)]
28. Sirbiladze, G.; Khvedelidze, T. Associated Statistical Parameters’ Aggregations in Interactive MADM. *Mathematics* **2023**, *11*, 1061. [[CrossRef](#)]
29. Sirbiladze, G.; Kacprzyk, J.; Manjafarashvili, T.; Midodashvili, B.; Matsaberidze, B. New Fuzzy Extensions on Binomial Distribution. *Axioms* **2022**, *11*, 220. [[CrossRef](#)]
30. Sirbiladze, G.; Kacprzyk, J.; Davitashvili, T.; Midodashvili, B. Associated Probabilities in Insufficient Expert Data Analysis. *Mathematics* **2024**, *12*, 518. [[CrossRef](#)]
31. de Campos, L.M.; Lamata, M.T.; Moral, S. Distances between fuzzy measures through associated probabilities: Some applications. *Fuzzy Sets Syst.* **1990**, *35*, 57–68. [[CrossRef](#)]
32. Sirbiladze, G.; Gachechiladze, T. Restored fuzzy measures in expert decision-making. *Inf. Sci.* **2005**, *169*, 71–95. [[CrossRef](#)]
33. Grabisch, M. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets Syst.* **1997**, *92*, 167–189. [[CrossRef](#)]
34. Wu, J.-Z.; Zhang, Q. 2-order additive fuzzy measure identification method based on diamond pairwise comparison and maximum entropy principle. *Fuzzy Optim. Decis. Mak.* **2010**, *9*, 435–453. [[CrossRef](#)]
35. Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Ann. Math. Stat.* **1967**, *38*, 325–339. [[CrossRef](#)]
36. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
37. Kandel, A. On the control and evaluation of uncertain processes. *IEEE Trans. Autom. Control* **1980**, *25*, 1182–1187. [[CrossRef](#)]
38. Dubois, D.; Prade, H. *Possibility Theory*; Plenum Press: New York, NY, USA, 1988.
39. Yager, R.R. On the Entropy of Fuzzy Measures. *IEEE Trans. Fuzzy Syst.* **2000**, *8*, 453–561. [[CrossRef](#)]
40. Huang, H. *A New Index for Measuring the Difference Between Two Probability Distributions*; CC-BY 4.0; Qeios: London, UK, 2024. [[CrossRef](#)]
41. Sirbiladze, G.; Midodashvili, B.; Midodashvili, L.; Siphashvili, D. About One Representation-Interpreter of a Monotone Measure. *J. Comput. Cogn. Eng.* **2021**, *1*, 13–24. [[CrossRef](#)]

42. Wu, J.-Z.; Beliakov, G. Nonadditivity Index Oriented Decision Preference Information Representation and Capacity Identification. *Econ. Comput. Econ. Cybern. Stud. Res.* **2020**, *54*, 281–297.
43. Huang, L.; Wu, J.-Z.; Beliakov, G. Multicriteria correlation preference information (MCCPI) with nonadditivity index for decision aiding. *J. Intell. Fuzzy Syst.* **2020**, *39*, 3441–3452. [[CrossRef](#)]
44. Wu, J.-Z.; Beliakov, G. Nonadditivity index and capacity identification method in the context of multicriteria decision making. *Inf. Sci.* **2018**, *467*, 398–406. [[CrossRef](#)]
45. Grabisch, M.; Kojadinovic, I.; Meyer, P. A review of methods for capacity identification in Choquet integral based multi-attribute utility theory. *Eur. J. Oper. Res.* **2008**, *186*, 766–785. [[CrossRef](#)]
46. Huang, L.; Wu, J.-Z.; Xi, R.-J. Nonadditivity Index Based Quasi-Random Generation of Capacities and Its Application in Comprehensive Decision Aiding. *Mathematics* **2020**, *8*, 301. [[CrossRef](#)]
47. Grabisch, M. Alternative Representations of Discrete Fuzzy Measures for Decision Making. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **1997**, *5*, 587–607. [[CrossRef](#)]
48. Javier, M.; Guillaume, S.; Pilar, B. k-maxitive fuzzy measures: A scalable approach 2 to model interactions. *Fuzzy Sets Syst.* **2017**, *324*, 33–48.
49. Javier, M.; Serge, G.; Tewfik, S.; Pilar, B. An algorithm for computing the generalized interaction index for k-maxitive fuzzy measures. *J. Intell. Fuzzy Syst.* **2020**, *38*, 1–11.
50. Ünver, M.; Özçelik, G.; Olgun, M. A fuzzy measure theoretical approach for multi criteria decision making problems containing sub-criteria. *J. Intell. Fuzzy Syst.* **2018**, *35*, 6461–6468. [[CrossRef](#)]
51. Buck, A.R.; Anderson, D.T.; Keller, J.M.; Wilkin, T.; Islam, M.A. A Weighted Matrix Visualization for Fuzzy Measures and Integrals. In Proceedings of the 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, UK, 19–24 July 2020; pp. 1–8. [[CrossRef](#)]
52. Abdullah, L.; Awang, N.A.; Othman, M. Application of Choquet Integral-Fuzzy Measures for Aggregating Customers' Satisfaction. *Adv. Fuzzy Syst.* **2021**, *2021*, 1–8. [[CrossRef](#)]
53. Zhang, M.; Cao, C. A 2-order Additive Fuzzy Measure Identification Method Based on Hesitant Fuzzy Linguistic Interaction Degree and Its Application in Credit Assessment. *Res. Sq.* **2021**, preprint. [[CrossRef](#)]
54. Beliakov, G.; Divakov, D. On representation of fuzzy measures for learning Choquet and Sugeno integrals. *Knowl. Based Syst.* **2020**, *189*, 105134. [[CrossRef](#)]
55. Grabisch, M. The representation of importance and interaction of features by fuzzy measures. *Pattern Recognit. Lett.* **1996**, *17*, 567–575. [[CrossRef](#)]
56. Roubens, M. Interaction between criteria and definition of weights in MCDA problems. In Proceedings of the 44th Meeting of the European Working Group “Multicriteria Aid for Decisions”, Brussels, Belgium, 10 March 1996.
57. Kojadinovic, I. Modeling interaction phenomena using fuzzy measures: On the notions of interaction and independence. *Fuzzy Sets Syst.* **2002**, *135*, 317–340. [[CrossRef](#)]
58. Marichal, J.L.; Roubens, M. Dependence between criteria and multiple criteria decision aid. In Proceedings of the 2nd International Workshop on Preferences and Decisions, Trento, Italy, 1–3 July 1998; pp. 69–75.
59. Beliakov, G.; Cabrerizo, F.J.; Herrera-Viedma, E.; Wu, J.-Z. Random generation of k-interactive capacities. *Fuzzy Sets Syst.* **2020**, *430*, 48–55. [[CrossRef](#)]
60. Beliakov, G.; Divakov, D. Aggregation with dependencies: Capacities and fuzzy integrals. *Fuzzy Sets Syst.* **2022**, *446*, 222–232. [[CrossRef](#)]
61. Sirbiladze, G.; Sikharulidze, A. Weighted fuzzy averages in fuzzy environment, Parts I, II. *Int. J. Uncertain. Fuzziness Knowl.* **2003**, *11*, 139–172. [[CrossRef](#)]
62. Sirbiladze, G.; Manjafarashvili, T. Connections between Campos-Bolanos and Murofushi–Sugeno Representations of a Fuzzy Measure. *Mathematics* **2022**, *10*, 516. [[CrossRef](#)]
63. Sirbiladze, G. *Extremal Fuzzy Dynamic Systems: Theory and Applications*; IFSR International Series on Systems Science and Engineering 28; Springer: New York, NY, USA; Heidelberg, Germany; Dordrecht, The Netherlands; London, UK, 2013.
64. Sirbiladze, G. Associated Probabilities' Aggregations in Interactive MADM for q-Rung Orthopair Fuzzy Discrimination Environment. *Int. J. Intell. Syst.* **2020**, *35*, 335–372. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.