Fuzzy measure and probability distributions: distorted probabilities

Yasuo Narukawa[1], Vicenç Torra[2]

[1] Toho Gakuen

3-1-10 Naka, Kunitachi, Tokyo, 186-0004 Japan

e-mail: `narukawa@d4.dion.ne.jp`

[2]Institut d'Investigació en Intel·ligència Artificial - CSIC

Campus UAB s/n, 08193 Bellaterra, Catalonia, Spain

e-mail: `vtorra@iiia.csic.es`,

`http://www.iiia.csic.es/~vtorra`

Abstract

This work studies fuzzy measures and their application to data modeling. We focus on the particular case when fuzzy measures are distorted probabilities. We analyze their properties and introduce a new family of measures (m-dimensional distorted probabilities). The work finishes with the application of two dimensional distorted probabilities to a modeling problem. Results of the application of such fuzzy measure to data modeling using Choquet integral are discussed.

Keywords: Fuzzy measures, Distorted probabilities, Choquet integral, Model determination.

1 Introduction

Sugeno [25] and Choquet [4] integrals are well-known aggregation operators for numerical data (see *e.g.*, [12], [26], or [36] for details). Using them, aggregation is defined as the integration of a function (the data to be aggregated) with respect to a fuzzy measure (or capacity).

However, although Sugeno and Choquet integrals are powerful operators, their practical application is more difficult due to the fact that they require a large number of parameters. In fact, fuzzy measures require $2^{|X|}$ parameters (where $|X|$ is the number of information sources or input variables) because they are functions from the power set of X (2^X) into the unit interval.

To make the application of these integrals easy, several approaches have been considered in the literature. We underline two of them. Namely, *restricted* fuzzy measures and learning methods from examples.

Restricted fuzzy measures are those measures that require less than $2^{|X|}$ parameters because they are constrained to satisfy additional properties than *general* or *unrestricted* fuzzy measures. This is *e.g.* the case of Sugeno λ-measures [25], $\perp$-decomposable measures (see *e.g.* [12]), k-additive fuzzy measures [10], p-symmetric [21], or hierarchically decomposable ones [29]. Among these measures, we would like to underline k-additive fuzzy measures. They define a family of measures (defined in terms of increasing values of k) and have analogies with the measures considered in this work. The similarities on both approaches are described in more detail in Section 2.

The use of learning (or parameter determination) methods are an alternative approach to deal with the complexity of fuzzy measures. In this case, the parameters are learnt from a set of examples (defined in terms of *(input, output)* pairs). Work has been done in this direction, probably being [27] the first published paper on this matter. Recent work in this area can be found in [16], [20] and [27], to name a few.

In this work we study distorted probabilities or, equivalently, fuzzy measures that can be represented in terms of a probability distribution on X and a non-decreasing function (a distortion function). The distorted probabilities were suggested in experimental psychology [5]. Descriptive models that incorporate transformed or distorted probability distributions have been proposed and studied in the field of Economics, *e.g.* [13, 17]. See also [1] for an acount of the use of such measures in game theory.

More precisely, in this paper we show that every fuzzy measure can be represented as the difference of two distorted probabilities and introduce some condition for a fuzzy measure to be a distorted probability. We also present a

2

generalization of distorted probabilities to define m dimensional ones. Roughly speaking, such measures are defined in terms of (i) a partition of X and one probability distribution over each partition element and (ii) a distortion function to combine such probability distributions. Then, we describe an approach to learn such measures from examples.

As it will be shown later, any fuzzy measure can be interpreted in terms of a m dimensional distorted probability for a particular value of m. Thus, fuzzy measures can be constructed in terms of probability distributions (a set of them) and a function that distorts and/or combines these distributions.

The structure of the paper is as follows. In Section 2 we motivate the interest of studying distorted probabilities. Then, in Section 3 we review some definitions that are used through the paper. In Section 4 we introduce some results that relate fuzzy measures and probability distributions. In particular, we characterize distorted probabilities. Then, in Section 5 we generalize the concept of distorted probabilities to m dimensional probabilities. This section also studies some properties. Finally, in Section 6 we describe an approach for learning two dimensional distorted probabilities and we apply it to a modeling problem consisting on 8 variables. The paper finishes in Section 7 with the conclusions.

2   Motivation

Distorted probabilities have proved to be useful and convenient fuzzy measures. It is known that they generalize Sugeno λ-measures and $\perp$-decomposable ones, the former being a type of fuzzy measure widely used in the literature *e.g.* [18]. Additionally, they have been used, in conjunction with Choquet integral, for data modeling (see *e.g.* the algorithm in [31] and [15]). In this setting, as well as when used with the Sugeno integral, distorted probabilities have a simple interpretation. A distorted probability on a set of sources X can be understood as an evaluation of the importance/reliability of the sources in X (the probability distribution function) and a fuzzy quantifier (or a weighting vector) that establishes the relative importance of the values themselves (instead of the importance of the sources). This latter element, that corresponds to

3

the distortion function permits to introduce compensation or optimism in the Choquet integral. See *e.g.* [28] for details. Note that a Choquet integral with distorted probabilities is equivalent to the WOWA operator [28] or to the OWA with importances [34].

Additionally, distorted probabilities can be seen as complementary to other families of fuzzy measures. This is due to the fact that they can offer, in some cases, a compact representation that others cannot offer.

For illustration, we have compared a distorted probability with the well-known family of k-additive fuzzy measures [10]. It is known that larger values for k yield to larger sets of fuzzy measures. This can be formalized denoting by $KAFM_{k_0,X}$ the set of all k_0-additive fuzzy measures. Then, it is known that

$$KAFM_{1,X} \subset KAFM_{2,X} \subset KAFM_{3,X} \cdots \subset KAFM_{|X|,X}$$

where $KAFM_{1,X}$ equals the set of additive measures and $KAFM_{|X|,X}$ equals the set of all fuzzy measures over X. In other words, $KAFM_{|X|,X}$ is the set of unconstrained fuzzy measures.

Naturally, given a k_0-additive fuzzy measure, the smaller is k_0 the less parameters are required to define such measure. Therefore, in general, the smaller the parameter, the simpler and more convenient for applications. This is so because users can better define the measure and/or grasp their meaning. Also, when the measure is learnt from examples, less examples are required. Nevertheless, it is clear that not all measures can be represented using small values of k and, thus, some of them require complex definitions (complex in terms of the number of parameters).

Now, we present an example where it is shown that a distorted probability requires a k-additive fuzzy measure with $k = |X|$ to be represented. This means, that such measure that has a simple representation with a distorted probability cannot have such simple representation when a k-additive fuzzy measure formalism is used.

Example 1. *Let us consider a distorted probability $\mu_{\mathbf{p},\mathbf{w}}$ over $X = \{x_1, x_2, x_3, x_4, x_5\}$ generated by the probability distribution $\mathbf{p} = (0.2, 0.3, 0.1, 0.2, 0.1)$ and a distortion function (or fuzzy quantifier) generated from the vector $\mathbf{w} = (0.1, 0.2, 0.4, 0.2, 0.1)$. The measure for all $Y \subseteq X$ is given in Table 1 (column*

$\mu_{\mathbf{p},\mathbf{w}}$).

Following the interpretation of WOWA and OWA with importances, when such measure $\mu_{\mathbf{p},\mathbf{w}}$ is used in a Choquet integral to aggregate $f(x_i)$, it corresponds to a situation in which the importance of x_i is p_i and in which the central values of $\{f(x_i)\}_i$ have 4 times the importance of the extreme ones and 2 times the importance of medium values. This is, the relevance of the values $\max f(x_i)$ and $\min f(x_i)$ is diminished, while the relevance of the central values is increased.

The Möbius transform of $\mu_{\mathbf{p},\mathbf{w}}$ is also given in Table 1. As the Möbius transform is different from zero for all $Y \subseteq X$ this means that $\mu_{\mathbf{p},\mathbf{w}}$ is a 5 additive fuzzy measure. This is, there is no k additive fuzzy measure for $k < 5$ equivalent to $\mu_{\mathbf{p},\mathbf{w}}$.

This example shows that a distorted probability can be in some circumstances an adequate alternative to k-additive fuzzy measures. Nevertheless, the modeling capabilities of such measures is limited. This is so because the *proportion* of fuzzy measures representable by distorted probabilities is rather small. The following example (from [14]) illustrates the case for $X = \{1, 2, 3\}$. It considers the possible orderings on 2^X of $\mu(Y)$ for $Y \subseteq X$ when a fuzzy measure is either a distorted probability or an unconstrained one. For simplicity, the example is restricted to measures such that $\mu(\{1\}) < \mu(\{2\}) < \mu(\{3\})$.

Example 2. Let μ be a distorted probability with $\mu(\{1\}) < \mu(\{2\}) < \mu(\{3\})$, then either:

$$\mu(\emptyset) < \mu(\{1\}) < \mu(\{2\}) < \mu(\{3\}) < \mu(\{1,2\}) < \mu(\{1,3\}) < \mu(\{2,3\}) < \mu(X)$$

or

$$\mu(\emptyset) < \mu(\{1\}) < \mu(\{2\}) < \mu(\{1,2\}) < \mu(\{3\}) < \mu(\{1,3\}) < \mu(\{2,3\}) < \mu(X)$$

holds.

Instead, if μ is an unconstrained fuzzy measure with $\mu(\{1\}) < \mu(\{2\}) < \mu(\{3\})$, then either:

$$\emptyset < \mu(\{1\}) < \mu(\{2\}) < \mu(\{1,2\}) < \mu(\{3\}) < \mu(\{1,3\}) < \mu(\{2,3\}) < \mu(X)$$

$$\emptyset < \mu(\{1\}) < \mu(\{2\}) < \mu(\{3\}) < \mu(\{1,2\}) < \mu(\{1,3\}) < \mu(\{2,3\}) < \mu(X)$$

$$\emptyset < \mu(\{1\}) < \mu(\{2\}) < \mu(\{3\}) < \mu(\{1,3\}) < \mu(\{1,2\}) < \mu(\{2,3\}) < \mu(X)$$

$$\emptyset < \mu(\{1\}) < \mu(\{2\}) < \mu(\{3\}) < \mu(\{1,3\}) < \mu(\{2,3\}) < \mu(\{1,2\}) < \mu(X)$$

$$\emptyset < \mu(\{1\}) < \mu(\{2\}) < \mu(\{1,2\}) < \mu(\{3\}) < \mu(\{2,3\}) < \mu(\{1,3\}) < \mu(X)$$

$$\emptyset < \mu(\{1\}) < \mu(\{2\}) < \mu(\{3\}) < \mu(\{1,2\}) < \mu(\{2,3\}) < \mu(\{1,3\}) < \mu(X)$$

$$\emptyset < \mu(\{1\}) < \mu(\{2\}) < \mu(\{3\}) < \mu(\{2,3\}) < \mu(\{1,2\}) < \mu(\{1,3\}) < \mu(X)$$

$$\emptyset < \mu(\{1\}) < \mu(\{2\}) < \mu(\{3\}) < \mu(\{2,3\}) < \mu(\{1,3\}) < \mu(\{1,2\}) < \mu(X)$$

Thus, fuzzy measures in a proportion of 2/8 are distorted probabilities. Table 2 gives the corresponding values for sets X of larger dimension. It can be seen that the proportion of distorted probabilities decreases rapidly when $|X|$ increases (*e.g.* 14 over 70016 when $|X| = 4$).

In this paper we propose m-dimensional distorted probabilities. This family, in a way similar to k-additive fuzzy measures, recovers for increasing values of m the space between $KAFM_{1,X}$ and $KAFM_{|X|,X}$. In other words, we try to make a smooth transition between the columns two and three of Table 2.

Although this family of fuzzy measures has resemblance with k-additive ones in the sense that they cover the whole space, as it has been said, the sets of fuzzy measures they define are different and thus they would be used in a complementary way in real applications. For a particular application, a particular family would be selected with care and on the basis of clarity and understandability.

3   Preliminaries

This section starts reviewing some basic definitions about fuzzy measures. In particular, besides of the definition of (unconstrained) fuzzy measures, we define distorted probabilities. Then, we review some results that are needed in the rest of the paper.

For the sake of simplicity, we will consider X to be a finite set and we suppose that $X := \{1, 2, \cdots, n\}$.

3.1 Definitions

Definition 1. *A function μ on $(X, 2^X)$ is a fuzzy measure if it satisfies the following axioms:*

(i) *$\mu(\emptyset) = 0$, $\mu(X) = 1$ (boundary conditions)*

(ii) *$A \subseteq B$ implies $\mu(A) \leq \mu(B)$ (monotonicity)*

We use also the term *unconstrained* fuzzy measure to refer to measures satisfying this definition. This is to distinguish them from other measures that satisfy other constraints besides conditions (i) and (ii). Distorted probabilities defined below are an example of *constrained* fuzzy measures.

Definition 2. *Let f be a real valued function on X and let P be a probability measure on $(X, 2^X)$. We say that f and P represent a fuzzy measure μ on $(X, 2^X)$ if and only if $\mu(A) = f(P(A))$ for all $A \in 2^X$.*

Definition 3. *Let f be a real valued function on X. We say that f is strictly increasing with respect to a probability measure P if and only if $P(A) < P(B)$ implies $f(P(A)) < f(P(B))$.*

Accordingly, μ is a distortion of a probability measure P when f is strictly increasing with respect to P. In general, when we know that there is such probability measure P and a strictly increasing function f for μ, we will say that μ is a distorted probability.

Definition 4. *Let μ be a fuzzy measure on $(X, 2^X)$. We say that μ is a distorted probability if it is represented by a probability distribution P on $(X, 2^X)$ and a function f that is strictly increasing with respect to a probability P.*

Remark. Since we suppose that X is a finite set, a strictly increasing function f with respect to P can be regarded as a strictly increasing function on $[0, 1]$ if there is no restriction of the function f. Points except $\{P(A) | A \in 2^X\}$ in $[0, 1]$ are not essential in this paper.

3.2 Previous results

In this section we review some propositions and theorems that are used latter on to give a representation theorem for distorted probabilities. First, we give a lemma that is equivalent to the unique decomposition for integers.

Lemma 5. *Suppose that for integers $0 < n_1 < n_2 < \cdots < n_k$ and $0 < m_1 < m_2 < \cdots < m_l$*

$$2^{n_1} + 2^{n_2} + \cdots + 2^{n_k} = 2^{m_1} + 2^{m_2} + \cdots + 2^{m_l}.$$

Then we have $k = l$ and $n_i = m_i$ for $i = 1, 2, \ldots, k$.

Now, we turn into some theorems obtained by Dana Scott in [24].

Definition 6. *Let L be a finite-dimensional real linear vector space, then a linear functional on L is a real-valued, homogeneous, additive function defined on L.*

Definition 7. *Let L be a finite-dimensional real linear vector space, then:*

1. *A subset $X \subseteq L$ is* symmetric *if $X = -X = \{-x : x \in X\}$*

2. *A subset $N \subseteq X$ is called* realizable *in X if there is a linear functional ϕ on L such that for all $x \in X$*

$$x \in N \text{ if and only if } \phi(x) \geq 0$$

When $N \neq X$, this latter condition means that there is a half-space H of L separating the sets N and $X \setminus N$ so that $X \cap H = N \cap H$. Instead, when $N = X$, the trivial 0 functional shows that N is realizable.

Following Scott [24], we use $x \succeq 0$ to mean $x \in N$, and $x \preceq 0$ to mean $-x \succeq 0$, and $x \succ 0$ for *not $x \preceq 0$.*

Theorem 8. *(Scott [24], Theorem 1.1) Let L be a finite-dimensional real linear vector space, let X be a finite, symmetric subset of L. For a subset $\{x \in X : x \succeq 0\}$ to be realizable in X it is necessary and sufficient that the conditions*

$$x \succeq 0 \quad \text{or} \quad x \preceq 0$$

$$\sum_{i \leq n} \lambda_i x_i = 0 \quad \text{implies} \quad x_1 \preceq 0$$

hold for all $x \in X$ and all sequences $x_1, \ldots, x_n \in X$, and all scalars $\lambda_1, \ldots, \lambda_n$ where $\lambda_i > 0$ and $x_i \succeq 0$, for $i \leq n$, and $n > 1$.

$\succeq$ is realizable if there exists a linear function ϕ on L such that for all $x, y \in X$ we have that $x \succeq y$ if and only if $\phi(x) \geq \phi(y)$.

Theorem 9. *(Scott [24], Theorem 4.1) Let B be a finite Boolean algebra and let $\succeq$ be a binary relation on B. For $\succeq$ to be realizable by a probability measure on B it is necessary and sufficient that the conditions:*

(1) $1 \succ 0$

(2) $x \succeq 0$

(3) $x \succeq y$ *or* $x \preceq y$

(4) $x_1 + x_2 + \cdots + x_n = y_1 + y_2 + \cdots + y_n$ *implies* $x_1 \preceq y_1$,

hold for all $x, y \in B$ and all sequences $x_1, \ldots, x_n, y_1, \ldots, y_n \in B$, where $x_i \succeq y_i$ for $i \leq n$, $i > 1$, and $n > 1$.

In this last theorem, $x_1 + x_2 + \cdots + x_n$ corresponds to the algebraic sum of characteristic functions and does not stand for the union of x_i. As [24] points out, this means that "every point (atom) belongs to exactly the same number of the x_i as the y_i".

4 Distorted probabilities

Now, we present some new results on distorted probabilities. In particular, we give a characterization of such measures.

4.1 Polynomial representation and distorted probabilities

Theorem 10. *For every fuzzy measure μ on $(X, 2^X)$, there exist a polynomial f and probability P on $(X, 2^X)$ such that $\mu = f \circ P$*

Proof. Define a probability measure P on $(X, 2^X)$ for all $k \in X$ by

$$P(\{k\}) := \frac{2^{k-1}}{2^n - 1}.$$

It follows from Lemma 5 that $P(A) \neq P(B)$ if $A \neq B$ for $A, B \in 2^X$. Define the $(2^n - 1) \times (2^n - 1)$ matrix by (assuming an ordering on 2^X)

$$\mathbf{P} = \begin{pmatrix} P(\{1\})^{2^n-1} & P(\{1\})^{2^n-2} & \ldots & P(\{1\}) \\ P(\{2\})^{2^n-1} & P(\{2\})^{2^n-2} & \ldots & P(\{2\}) \\ \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \\ P(X)^{2^n-1} & P(X)^{2^n-2} & \ldots & P(X) \end{pmatrix}.$$

Since $P(A) \neq P(B)$ for $A \neq B, A, B \in 2^X$, we have $rank\mathbf{P} = 2^n - 1$. Therefore there exists the inverse $\mathbf{P}^{-1}$ of $\mathbf{P}$. Denote

$$
\begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_{2^n-1} \end{pmatrix} := \mathbf{P}^{-1} \begin{pmatrix} \mu(\{1\}) \\ \mu(\{2\}) \\ \cdot \\ \cdot \\ \cdot \\ \mu(X) \end{pmatrix}.
$$

Then define the polynomial by

$$
f(x) := a_1 x^{2^n-1} + a_2 x^{2^n-2} + \cdots + a_{2^n-1} x.
$$

This is the polynomial which represents the fuzzy measure μ. Indeed, we have $f(0) = 0$ and $\mu(A) = f(P(A))$ for all $A \in 2^X$ from the definition of coefficients a_k $(k = 1, 2, \ldots, 2^n - 1)$. $\qquad\square$

Define b_k by $b_k := a_k \vee 0$, c_k by $c_k := -(a_k \wedge 0)$. Denote

$$
f^+(x) := \sum_k b_k x^{2^n-k}
$$

and

$$
f^-(x) := \sum_k c_k x^{2^n-k}.
$$

Then f^+ and f^- are strictly increasing. Therefore, we have the next corollary.

Corollary 11. *For every fuzzy measure μ on $(X, 2^X)$, there exist strictly increasing polynomials f^+, f^- and a probability P on $(X, 2^X)$ such that*

$$
\mu = f^+ \circ P - f^- \circ P. \tag{1}
$$

Therefore, any fuzzy measure can be expressed as the difference of two distorted probabilities.

If $|X| = 2$, every fuzzy measure on $(X, 2^X)$ is a distorted probability. This is proven in the following corollary.

Corollary 12. *Let μ be a fuzzy measure on $(X, 2^X)$ with $X = \{1, 2\}$ and, without loss of generality, $1 > \mu(\{2\}) \geq \mu(\{1\}) > 0$. Then, there exists a strictly increasing function f and a probability P that represent a fuzzy measure μ.*

10

Proof. If $0 < \mu(\{1\}) < \mu(\{2\}) < 1$, then define a probability P as in Theorem 10, that is,

$$P(\{1\}) := \frac{1}{3}, P(\{2\}) := \frac{2}{3}.$$

Then it follows from Theorem 10 that there exists a polynomial f that represents μ. It follows from the definition of P that $P(\{1\}) < P(\{2\}) < P(\{1,2\}) = 1$. This means that $f(P(\{1\})) < f(P(\{2\})) < f(P(\{1,2\})) = 1$. If $0 < \mu(\{1\}) = \mu(\{2\}) < 1$, then define a probability P by

$$P(\{1\}) = P(\{2\}) := \frac{1}{2}.$$

Then, it is obvious that there exists a strictly increasing polynomial f that represents μ. $\square$

However, neither any polynomial f in Theorem 10 nor any pair of polynomials f^+ and f^- leads to fuzzy measures. Only non-decreasing polynomials f in Theorem 10 are assured to lead to them.

4.2 Characterization of distorted probabilities

Definition 13. *Let μ be a fuzzy measure on $(X, 2^X)$. If $\mu(A) < \mu(B) \Leftrightarrow \mu(A \cup C) < \mu(B \cup C)$ for every $A \cap C = \emptyset$, $B \cap C = \emptyset$ $A, B, C \in 2^X$, we say that μ is a pre-distorted probability.*

Proposition 14. *Suppose that a function f and a probability P represent a fuzzy measure μ.*

If f is strictly increasing with respect to P, then $\mu(A) < \mu(B) \Leftrightarrow \mu(A \cup C) < \mu(B \cup C)$ for every $A \cap C = \emptyset$, $B \cap C = \emptyset$ $A, B, C \in 2^X$. In other words, distorted probabilities are pre-distorted probabilities.

If f is non-decreasing with respect to P, then $\mu(A) < \mu(B)$ implies $\mu(A \cup C) \le \mu(B \cup C)$ for every $A \cap C = \emptyset$, $B \cap C = \emptyset$ $A, B, C \in 2^X$.

Proof. Suppose that $\mu(A) < \mu(B)$ for $A, B \in 2^X$ and $\mu = f \circ P$. If f is strictly increasing, we have $P(A) < P(B)$. Then for every $A \cap C = \emptyset$ and $B \cap C = \emptyset$, we have

$$P(A \cup C) = P(A) + P(C) < P(B) + P(C) = P(B \cup C).$$

It follows from the fact that f is strictly increasing that $f \circ P(A \cup C) < f \circ P(B \cup C)$, that is, $\mu(A \cup C) < \mu(B \cup C)$. It is much the same to prove $\mu(A \cup C) < \mu(B \cup C) \Rightarrow \mu(A) < \mu(B)$.

When f is non-decreasing, we have $P(A) < P(B)$. Then for every $A \cap C = \emptyset$ and $B \cap C = \emptyset$, we have

$$P(A \cup C) = P(A) + P(C) < P(B) + P(C) = P(B \cup C).$$

It follows from the fact that f is non-decreasing that $f \circ P(A \cup C) \leq f \circ P(B \cup C)$, that is, $\mu(A \cup C) \leq \mu(B \cup C)$. $\square$

Proposition 15. *Let μ be a pre-distorted probability on $(X, 2^X)$. If $\mu(A) < \mu(B)$, then $\mu(A^c) > \mu(B^c)$ for $A, B \in 2^X$.*

Proof. Let us consider two cases. First, we suppose that $A \cap B = \emptyset$. Then, since μ is a pre-distorted probability, we have $A \subset B^c$ and $B \subset A^c$. Therefore it follows from $A^c = B \cup (A^c \cap B^c)$ and $B^c = A \cup (B^c \cap A^c)$ that $\mu(A^c) > \mu(B^c)$.

Second, we suppose that $A \cap B = D$. Then, $\mu(A) < \mu(B)$ if and only if $\mu(A') < \mu(B')$ where $A' = A \setminus D$ and $B' = B \setminus D$. Therefore, using the case above $\mu(A'^c) > \mu(B'^c)$. Naturally, $A'^c = A^c \cup D$ and $B'^c = B^c \cup D$. Therefore, as μ is a pre-distorted probability, $\mu(A'^c) > \mu(B'^c)$ if and only if $\mu(A^c) > \mu(B^c)$. $\square$

We consider a condition that plays a central role in the characterization of distorted and pre-distorted probabilities.

Definition 16. *Let μ be a fuzzy measure on $(X, 2^X)$, μ satisfies* Scott's condition *when for all $A_i, B_i \in 2^X$ condition (i) and condition (ii) below hold:*

(i) $\sum_{i=1}^{n} 1_{A_i} = \sum_{i=1}^{n} 1_{B_i}$

(ii) $\mu(A_i) \leq \mu(B_i)$ *for $i = 2, 3, \ldots, n$ implies $\mu(A_1) \geq \mu(B_1)$*

Now, we give a characterization of distorted probabilities in terms of Scott's condition.

Theorem 17. *Let μ be a fuzzy measure on $(X, 2^X)$, then μ is a distorted probability if and only if Scott's condition holds.*

Proof. First, we prove that there exists a probability P such that $P(A) \leq P(B)$ if and only if $\mu(A) \leq \mu(B)$. To do so, let us define the binary relation $\preceq$ on 2^X by

$$A \preceq B \Leftrightarrow \mu(A) \leq \mu(B)$$

for A and $B \in 2^X$. It is obvious from this definition that $\emptyset \preceq A$ and $A \preceq B$ for all $A, B \in 2^X$ such that $A \subseteq B$. Then, we have $\emptyset \prec X$ where $A \prec B$ means $A \preceq B$ and not $B \preceq A$. Also, from Scott's condition it follows that $\sum_{i=1}^{n} 1_{A_i} = \sum_{i=1}^{n} 1_{B_i}$ and $A_i \preceq B_i$ for $i = 2, 3, \ldots, n$ implies $B_1 \succeq A_1$ for $A_i, B_i \in 2^X$. Therefore, conditions of Theorem 9 (Scott [24], Theorem 4.1) are satisfied and, thus, there exists a probability measure P such that $A \preceq B$ if and only if $P(A) \leq P(B)$. So, according to $\preceq$ definition, $\mu(A) \leq \mu(B)$ if and only if $P(A) \leq P(B)$.

Now, let us assume that

$$0 < \mu(A_1) < \mu(A_2) < \cdots < \mu(A_{k-1}) < \mu(X) = 1$$

$$0 < P(A_1) < P(A_2) < \cdots < P(A_{k-1}) < P(X) = 1$$

where k $(k \leq (2^n - 2))$ is the maximum number so that the inequalities above hold.

Then, define the following $k \times k$ matrix by

$$\mathbf{P} = \begin{pmatrix} P(A_1)^k & P(A_1)^{k-1} & \ldots & P(A_1) \\ P(A_2)^k & P(A_2)^{k-1} & \ldots & P(A_2) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ P(X)^k & P(X)^{k-1} & \ldots & P(X) \end{pmatrix},$$

Since $P(A_l) < P(B_m)$ for $l < m$, we have $rank\mathbf{P} = k$. Therefore there exists the inverse $\mathbf{P}^{-1}$ of $\mathbf{P}$. Following the same approach considered in the proof of Theorem 10, we define the polynomial

$$f(x) := a_1 x^k + a_2 x^{k-1} + \cdots + a_k x.$$

where $(a_1 \ldots a_k)^T = \mathbf{P}^{-1} (\mu(A_1) \ldots \mu(A_k))^T$.

This polynomial represents the fuzzy measure μ. Indeed, we have $f(0) = 0$ and $\mu(A) = f(P(A))$ for all $A \in 2^X$.

13

Since $P(A) < P(B)$ implies $\mu(A) < \mu(B)$, f is strictly increasing with respect to P.

$\square$

Chateauneuf [3] obtains results similar to Theorem 17 using the results by Fishburn in [7]. In Theorem 17, the distortion function f is a strictly increasing polynomial while Chateauneuf only presents necessary and sufficient conditions for the existence of a non-decreasing distortion function.

4.3   Distorted and pre-distorted probabilities

Theorem 17 characterizes distorted probabilities. This theorem, together with Proposition 14 implies the following proposition.

Proposition 18. *Let μ be a fuzzy measure on $(X, 2^X)$ satisfying Scott's condition, then μ is a pre-distorted probability.*

The reversal of Proposition 18 is not true. Not all pre-distorted probabilities are distorted probabilities. This can be seen in the next example.

Example 3. *Let $X := \{1, 2, 3\}$, and let μ be the fuzzy measure on 2^X defined as:*

$$\mu(\{1\}) := \frac{1}{7}, \mu(\{2\}) := 0, \mu(\{3\}) := \frac{2}{7}, \mu(\{1,2\}) := \frac{3}{7}, \mu(\{2,3\}) := \frac{4}{7}, \mu(\{1,3\}) := 1.$$

Then, this measure is pre-distorted probability because:

$$\mu(\{2\}) < \mu(\{1\}), \mu(\{2,3\}) < \mu(\{1,3\})$$

$$\mu(\{2\}) < \mu(\{3\}), \mu(\{1,2\}) < \mu(\{1,3\})$$

$$\mu(\{1\}) < \mu(\{3\}), \mu(\{1,2\}) < \mu(\{2,3\}).$$

However, this measure does not satisfy Scott's condition. To illustrate this, let us consider the following sets: $A_1 := \{1\}, A_2 := \{2\}, A_3 := \{3\}, A_4 := X$ and $B_1 := \{1,2\}, B_2 := \emptyset, B_3 := \{2,3\}, B_4 := \{1,3\}$. Then, it holds that

$$1_{A_1} + 1_{A_2} + 1_{A_3} + 1_{A_4} = 1_{B_1} + 1_{B_2} + 1_{B_3} + 1_{B_4}$$

and

$$\mu(A_2) \leq \mu(B_2), \mu(A_3) \leq \mu(B_3), \mu(A_4) \leq \mu(B_4)$$

However,

$$\mu(A_1) < \mu(B_1).$$

Example 3 together with Theorem 17 illustrate that not all pre-distorted probabilities are distorted probabilities. Nevertheless, although the measure in Example 3 can not be represented using a strictly increasing distortion function f it is representable using a non-decreasing f. The corresponding representation is given in the following example.

Example 4. *Let us consider X and the measure μ defined in Example 3, then, there exists probability distributions P and non-decreasing functions f such that $\mu = f \circ P$. Table 3 gives one of such probability distributions together with the fuzzy measure μ.*

Example 4 shows that there exists a non-decreasing function f on $\{A | A \in 2^X\}$ and a probability distribution P such that $\mu = f \circ P$, but such f is no longer a polynomial. In fact, since f is non-decreasing with $f(0) = f(1/5.5) = 0$, there exist infinitely many a such that $f(a) = 0$. Therefore, because of Gauss' fundamental theorem of algebra, a polynomial of this form is impossible.

Let μ be a fuzzy measure on $(X, 2^X)$. We define the equivalence relation $\sim$ on 2^X by

$$A \sim B \Leftrightarrow \mu(A) = \mu(B)$$

for $A, B \in 2^X$. Note that this is the symmetric part of relation $\succeq$ in Theorem 17. Then, let us consider the quotient set $2^X / \sim$ (*i.e.*, if $[A] \in 2^X / \sim$, $B \in [A]$ is equivalent to $A \sim B$) and let $\mathcal{B}_\mu$ denote its representatives ($\emptyset$ and X are considered in $\mathcal{B}_\mu$). Naturally, $A, B \in \mathcal{B}_\mu$ implies either $\mu(A) < \mu(B)$ or $\mu(B) < \mu(A)$. Let L be the real linear vector space generated by the set of characteristic functions $1_A : A \in 2^X$ and let $\mathcal{X}_\mu$ of L be defined by

$$\mathcal{X}_\mu := \{1_A - 1_B | A, B \in \mathcal{B}_\mu\}.$$

The function $f : [0, 1] \to [0, 1]$ is said to be strictly increasing with respect to $\mathcal{B}_\mu$ if and only if $P(A) < P(B)$ implies $f(P(A)) < f(P(B))$ for $A, B \in \mathcal{B}_\mu$.

Now, we define *Scott's twin condition*. This condition is analogous to Scott's condition (Definition 16) but restricted to sets in $\mathcal{B} \subset 2^X$.

Definition 19. *Let μ be a fuzzy measure on $(X, 2^X)$ and let $\mathcal{B}$ be a subset of 2^X, then μ satisfies* Scott's twin condition *when for all $A_i, B_i \in \mathcal{B}$ conditions (i) and (ii) below hold:*

(i) $\sum_{i=1}^{n} 1_{A_i} = \sum_{i=1}^{n} 1_{B_i}$

(ii) $\mu(A_i) \le \mu(B_i)$ *for* $i = 2, 3, \ldots, n$ *implies* $\mu(A_1) \ge \mu(B_1)$

Suppose that $A, B \in \mathcal{B}_\mu$ and $\mu(A) \le \mu(B)$. It follows from the definition of $\mathcal{B}_\mu$ that the equality occur if and only if $A = B$.

Theorem 20. *Let μ be a fuzzy measure on $(X, 2^X)$. There exists a probability P on $(X, 2^X)$ and a polynomial f which is strictly increasing with respect to $\mathcal{B}_\mu$ such that*

$$\mu = f \circ P$$

if and only if Scott's twin condition holds for $\mathcal{B}_\mu$.

Proof. Let L be the real linear vector space generated by the set of characteristic functions 1_A for all $A \in 2^X$. Since $\emptyset \in \mathcal{B}_\mu$, $-1_A = 1_\emptyset - 1_A \in \mathcal{X}_\mu$. Therefore $-\mathcal{X}_\mu := \{-f \mid f \in \mathcal{X}_\mu\} = \mathcal{X}_\mu$, that is, $\mathcal{X}_\mu$ is a symmetric set. Then, let us define the relation $0 \preceq 1_A - 1_B$ on $\mathcal{X}_\mu$ is defined by $0 \preceq 1_A - 1_B \Leftrightarrow \mu(A) \le \mu(B)$ for $A, B \in \mathcal{B}_\mu$. Then we have that for $x \in \mathcal{X}_\mu$, $x \preceq 0$ means $0 \preceq (-x)$ and $0 \prec x$ means that it is not true $x \preceq 0$.

Suppose that Scott's twin condition holds. Then, we have

$$\sum_{i=0}^{n} x_i = 0 \text{ and } 0 \prec x_i \text{ for } i \ge 1$$

implies $x_0 \prec 0$ for $x_i \in \mathcal{X}_\mu$, $(0 \le i \le n)$. Since the conditions of Theorem 8 (Scott [24], Theorem 1.1) are satisfied, there exists a linear function ϕ on L such that for all $x \in \mathcal{X}_\mu$ $0 \prec x$ if and only if $\phi(x) > 0$. Since $\phi(1_X) > 0$, we can define a probability on 2^X by

$$P(A) := \frac{\phi(1_A)}{\phi(1_X)}.$$

Since ϕ is linear, $P(A) < P(B) \Leftrightarrow \phi(1_B - 1_A) > 0 \Leftrightarrow 1_A \prec 1_B \Leftrightarrow \mu(A) < \mu(B)$ for $A, B \in \mathcal{B}_\mu$. We may suppose the probabilities of the representatives A of $\mathcal{B}_\mu$ are maximal with respect to $[A]$, that is, satisfying $P(A) \ge P(A')$ for $A' \in [A]$, where $[A]$ is $[A] \in 2^X / \sim$, $A \in [A]$. In the same way as Theorem 18 we can show

16

that there exists a strictly increasing polynomial f such that $\mu(A) = f \circ P(A)$ for $A \in \mathcal{B}_\mu$. $\square$

Remark. Since $\mathcal{B}_\mu$ is not Boolean algebra, we cannot apply Theorem 9 directly.

Corollary 21. *Let μ be a fuzzy measure on $(X, 2^X)$. If μ is represented by a probability P on $(X, 2^X)$ and a non-decreasing function f, then Scott's twin condition holds.*

Proof. Let us define $\mathcal{B}_\mu$ as above. Then, suppose that $\mu(A) < \mu(B)$ for $A, B \in \mathcal{B}_\mu$. Since f is non-decreasing we have $P(A) < P(B)$. Conversely if $P(A) < P(B)$ for $A, B \in \mathcal{B}_\mu$, it follows from the definition of $\mathcal{B}_\mu$ that $\mu(A) < \mu(B)$. Then there exists a polynomial f^* that is strictly increasing with respect to $\mathcal{B}_\mu$ such that $\mu(A) = f^* \circ P(A)$ for $B \in \mathcal{B}_\mu$. Therefore it follows from Theorem 20 that Scott's twin condition holds. $\square$

Example 5. *Let $X := \{1, 2, 3\}$, $\mu(\emptyset) = \mu(\{1\}) = \mu(\{2\}) = 0$, $\mu(\{3\}) = 0.5$ $\mu(\{1, 2\}) = 0.7$, $\mu(\{1, 3\}) = \mu(\{2, 3\}) = \mu(X) = 1$. Then $\mathcal{B}_\mu = \{\emptyset, \{3\}, \{1, 2\}, X\}$. Since we have*

$$1_{\{1,2\}} + 1_{\{3\}} = 1_\emptyset + 1_X,$$

$\mu(\{1, 2\}) < \mu(X)$ and $\mu(\{3\}) > \mu(\emptyset)$. Therefore there exist a probability P and a strictly increasing polynomial with respect to $\mathcal{B}_\mu$ such that $\mu = f \circ P$

Example 6. *Let $X := \{1, 2, 3, 4\}$, $\mu(\emptyset) = \mu(\{1\}) = \mu(\{2\}) = \mu(\{3\}) = \mu(\{4\}) = 0$, $\mu(\{1, 2\}) = 0.3$, $\mu(\{2, 3\}) = 0.2$, $\mu(\{1, 4\}) = 0.1$, $\mu(\{3, 4\}) = 0.4$, and $\mu(A) = 1$ otherwise. This fuzzy measure cannot be representable in terms of a distortion function. There is no strictly increasing polynomial w.r.t. $\mathcal{B}_\mu$ and a probability P such that $\mu = f \circ P$. This is so because we have*

$$\mathcal{B}_\mu := \{\emptyset, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 4\}, X\},$$

such that

$$1_{\{1,2\}} + 1_{\{3,4\}} = 1_{\{1,4\}} + 1_{\{2,3\}},$$

with $\mu(\{1, 2\}) > \mu(\{2, 3\})$ and $\mu(\{3, 4\}) > \mu(\{1, 4\})$.

Definition 22. *Let $a_1, a_2, \ldots, a_m \in R^n$. The convex cone $[a_1, a_2, \ldots, a_m]$ generated by $\{a_1, a_2, \ldots, a_m\}$ is defined by*

$$[a_1, a_2, \ldots, a_m] := \{\sum_{i=1}^{m} \lambda_i a_i | \lambda_i \geq 0\}.$$

Theorem 23. *Let μ be a fuzzy measure on $(X, 2^X)$ with $\mathcal{B}_\mu = 2^X$. If there exist linearly independent vectors $a_1, a_2, \ldots, a_m \in R^n$ such that $\mathcal{X}_\mu \subset [a_1, a_2, \ldots, a_m] \backslash 0$, then μ is a distorted probability.*

Proof. Since $L := [a_1, a_2, \ldots, a_m] - [a_1, a_2, \ldots, a_m]$ is a linear space, we can define the linear function $\varphi(x)$ on L by

$$\varphi(x) := \sum_i \lambda_i$$

for $x = \sum_i \lambda_i a_i$. Since $\{a_1, a_2, \ldots, a_m\}$ is linearly independent, φ is well defined. Since $1_X \in \mathcal{X}_\mu$, there exists λ_i such that $\lambda_i > 0$. It follows from $\mathcal{B}_\mu = 2^X$ that we can define probability on $(X, 2^X)$ by $P(A) := \varphi(1_A)/\varphi(1_X)$ for $A \in 2^X$.

If $\mu(A) < \mu(B)$, $1_B - 1_A \in \mathcal{X}_\mu$. Therefore $\varphi(1_B - 1_A) > 0$, that is, $P(A) < P(B)$. $\square$

Corollary 24. *Let μ be a pre-distorted probability on $(X, 2^X)$. If $X = \{1, 2, 3\}$ and $\mathcal{B}_\mu = 2^X$, then μ is a distorted probability.*

Proof. Since $\mathcal{B}_\mu = 2^X$, we may suppose that $\mu(1) < \mu(2) < \mu(3)$. Since μ is pre-distorted probability, we have $\mu(\{1, 2\}) < \mu(\{1, 3\}) < \mu(\{2, 3\})$. Suppose that $\mu(\{1, 2\}) < \mu(3)$. Since $\mathcal{B}_\mu = 2^X$,

$$\mathcal{X}_\mu = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

where $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \in \mathcal{X}_\mu$ means that $1_{\{x_3\}} - 1_{\{x_1, x_2\}} \succ 0$.

Denote $a_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $a_2 := \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $a_3 := \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$. We have

$$\mathcal{X}_\mu \subset [a_1, a_2, a_3] \setminus 0.$$

18

Suppose that $\mu(\{1,2\}) > \mu(3)$. Since $\mathcal{B}_\mu = 2^X$,

$$\mathcal{X}_\mu = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

Denote $a_1 := \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$, $a_2 := \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$ and $a_3 := \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$. We have

$$\mathcal{X}_\mu \subset [a_1, a_2, a_3] \setminus 0.$$

$\square$

Corollary 25. *Let μ be a pre-distorted probability on $(X, 2^X)$. If $X = \{1, 2, 3, 4\}$ and $\mathcal{B}_\mu = 2^X$, then μ is a distorted probability.*

Proof. We may suppose that

$$0 < \mu(\{1\}) < \mu(\{2\}) < \mu(\{3\}) < \mu(\{4\}).$$

Since μ is a pre-distorted probability, we have

$$\mu(\{1,2\}) < \mu(\{1,3\}) < \mu(\{1,4\}).$$

It follows from Proposition 15 that

$$\mu(\{2,3\}) < \mu(\{2,4\}) < \mu(\{3,4\}),$$

$$\mu(\{1,2,3\}) < \mu(\{1,3,4\}) < \mu(\{2,3,4\}) < 1.$$

Suppose that $\mu(\{4\}) < \mu(\{1,2\})$ and $\mu(\{1,4\}) < \mu(\{2,3\})$. Denote $a_1 :=$
$\begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix}$, $a_2 := \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$, $a_3 := \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}$ and $a_4 := \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}$. Then we have

$$\mathcal{X}_\mu \subset [a_1, a_2, a_3, a_4] \setminus \{0\}.$$

In the other cases similarly we can define $a_1, a_2, a_3, a_4 \in R^n$ such that

$$\mathcal{X}_\mu \subset [a_1, a_2, a_3, a_4] \setminus \{0\}.$$

$\square$

If $|X| = 5$, there exists a pre-distorted probability μ such that μ is not distorted probability. The next example is based on [19].

Example 7. *Let $X := \{1, 2, 3, 4, 5\}$. Define a fuzzy measure according to the tables below.*

It is easy to check that $\mathcal{B}_\mu = 2^X$ and μ is pre-distorted probability.

Suppose that there exists a strictly increasing function f and Probability P such that $\mu = f \circ P$. Since f is non-decreasing, it follows from

$$\mu(\{2,3\}) < \mu(\{1\}), \mu(\{1,2\}) < \mu(\{3,4\}) \, and \, \mu(\{1,3\}) < \mu(\{2,5\})$$

that

$$P(\{2,3\}) < P(\{1\}), P(\{1,2\}) < P(\{3,4\}) \, and \, P(\{1,3\}) < P(\{2,5\}).$$

Then it follows from the additivity of P that $P(\{1,2,4\}) < P(\{3,5\})$. Therefore $f \circ P(\{1,2,4\}) \leq f \circ P(\{3,5\})$, that is $\mu(\{1,2,4\}) \leq \mu(\{3,5\})$. But we have $\mu(\{1,2,4\}) > \mu(\{3,5\})$ in the table.

Remark *The example above shows that μ satisfies the condition: $\mu(A) \leq \mu(B) \Rightarrow \mu(A \cup C) \leq \mu(B \cup C)$ for every $A \cap C = \emptyset, B \cap C = \emptyset, A, B, C \in 2^X$, and there exists no non-decreasing function f such that $\mu = f \circ P$ where P is probability on $(X, 2^X)$.*

5 m dimensional distorted probabilities

In this section, we introduce m dimensional distorted probabilities and give some properties.

Definition 26. *Let $\{X_1, X_2, \cdots, X_m\}$ be a partition of X (i.e., $\cup X_i = X$ and for all X_i and X_j it holds $X_i \cap X_j = \emptyset$), then we say that μ is at most m dimensional pre-distorted probability if for all X_i it holds:*

$$\mu(A) \geq \mu(B) \Leftrightarrow \mu(A \cup C) \geq \mu(B \cup C) \qquad (2)$$

for all $A, B, C \subseteq X_i$ such that $C \cap A = \emptyset$ and $C \cap B = \emptyset$. We say that at most m dimensional pre-distorted probability μ is m dimensional pre-distorted probability if μ is not at most $m - 1$ dimensional.

From the above definition for pre-distorted probabilities, it is clear that:

Proposition 27. *All fuzzy measures are at most $|X|$ dimensional pre-distorted probabilities.*

Definition 28. *Let $\{X_1, X_2, \cdots, X_m\}$ be a partition of X, then we say that μ is at most m dimensional distorted probability if there exists a function f on R^m and probability P_i on $(X_i, 2^{X_i})$ such that*

$$\mu(A) = f(P_1(A \cap X_1), P_2(A \cap X_2), \cdots, P_m(A \cap X_m)) \tag{3}$$

where f is strictly increasing with respect to the i-th axis for all $i = 1, 2, \ldots, m$. We say that at most m dimensional distorted probability μ is m dimensional pre-distorted probability if μ is not at most $m - 1$ dimensional.

From the above definition for distorted probabilities, we have the next proposition.

Proposition 29. *All fuzzy measures are at most $|X|$ dimensional distorted probabilities.*

Proof. Define a function $g : 2^X \to \{0, 1\}^n$ by $g(A) = (x_{A1}, x_{A2}, \ldots, x_{An})$, $x_{Ai} = 1$ if $x_i \in A$ and $x_{Ai} = 0$ if $x_i \notin A$ for $i = 1, 2, \ldots, n$. Then it is obvious that g is a bijection. Therefore we can define the function $f : \{0, 1\}^n \to R$ by $f(g(A)) := \mu(A)$. Define probability P_i on $\{x_i\}$ by $P_i(x_i) = 1$. Then we have $\mu(A) = f(P_1(A \cap \{x_1\}), \ldots, P_n(A \cap \{x_n\}))$.
$\square$

Example 8. *The Grabisch fuzzy measure on $X = \{M, L, P\}$ defined in [9] (and reproduced in Table 4) is both a two dimensional pre-distorted probability and at most two dimensional distorted probability. For building the distorted probability, we need to consider two sets. One set corresponds to the science subjects $\{M, P\}$ and the other corresponds to the literary subject $\{L\}$. A graphical interpretation of the measure is given in Figure 1. The figure represents the measure considering two axes. The axes are to represent the partition elements. Therefore, one axis corresponds to the set $\{L\}$ and the other to $\{M, P\}$. The figures represent the values of the measure. I.e., the figure for $\{L\} \times \{M, P\}$*

{L}	0.3	0.9	0.9	1.0
∅	0	0.45	0.45	0.5
	∅	{M}	{P}	{M,P}

Figure 1: Graphical interpretation of Grabisch fuzzy measure as a two dimensional distorted probability

corresponds to $\mu(\{L\} \cup \{M, P\}) = \mu(X) = 1$. It can be seen that the measure is increasing in both axes. Using $P_1(\{L\}) = 1$ and $P_2(\{M\}) = P_2(\{P\}) = 0.5$ and f according to Figure 1, we have $\mu(A) = f(P_1(A \cap \{L\}), P_2(A \cap \{M, P\}))$.

It is very important for both theory and applications to know the dimension of a distorted probability. Although the definitive method is still an open problem, using pre-distorted probability, we have the following results about the dimension of distorted probabilities.

Definition 30. *[22] Let $\mathcal{P} := \{X_1, X_2, \cdots, X_m\}$ be a partition of X. We say that $\mathcal{P}$ is an interadditive partition of X with respect to μ if*

$$\mu(A) = \sum_{i=1}^{n} \mu(A \cap X_i).$$

Proposition 31. *Let μ be m dimensional pre-distorted probability on $(X, 2^X)$ and $\{X_1, X_2, \cdots, X_m\}$ be an interadditive partition of X.*

If $|X_i| \leq 4$ for all X_i, μ is at most m dimensional distorted probability.

Proof. Let $g : R^m \to R$ be $g(x_1, \ldots, x_m) = x_1 + \cdots + x_m$. Since $\{X_1, X_2, \cdots, X_m\}$ is an interadditive partition, We have $\mu(A) = g(\mu(A \cap X_1), \ldots, \mu(A \cap X_m)))$. Applying Corollary 25, there exists a strictly increasing polynomial f_i and a probability P_i such that $\mu(A \cap X_i) = f_i(P_i(A \cap X_i))$ for $i = 1, \ldots, m$. Then define the function $f : R^m \to R$ by $f(x_1, \ldots, x_m) := g(f_1(x_1), \ldots, f_m(x_m))$. Therefore we have $\mu(A) := f(P(A \cap X_1), \ldots, P(A \cap X_m))$. Since f_i is strictly increasing, f is strictly increasing with respect to the i-th axis for all $i = 1, 2, \ldots, m$. $\qquad\square$

Corollary 32. *Let μ be a 2 dimensional pre-distorted probability on $(X, 2^X)$ and $\{X_1, X_2\}$ an interadditive partition of X. If $|X| = 5$, μ is at most 2 dimensional distorted probability.*

Example 9. *Let $X := \{1, 2, 3, 4, 5, 6, 7, 8\}$, $X_1 := \{1, 3, 5, 7\}$ and $X_2 := \{2, 4, 6, 8\}$. Then, a 2-dimensional pre-distorted probability is at most 2-dimensional distorted probability.*

6 Learning two dimensional distorted probability measures

This section is devoted to the learning of two dimensional distorted probabilities for data modeling using Choquet integrals. Our approach is based on the algorithm developed in [31] for learning distorted probabilities. Such algorithm is in turn based on the results described in [6] and [30, 32] for learning OWA [35] and weighted mean weights. The approach we follow is similar in nature to Tanaka and Murofushi [27], Grabisch [8], Marichal and Roubens [20] and Imai et al. [16] to name a few. These works deal with the process of fuzzy measure determination for Choquet integral (see [11] and [33] for reviews on this subject).

In all these cases, the learning process assumes a set of M examples consisting each one of N input values and the corresponding output value. Table 5 represents such examples. Each row $(a_1^i, a_2^i, \ldots, a_N^i | b^i)$ is an example, with a_j^i denoting the value supplied by the x_j information source in the i-th example and with b^i denoting the output value also for the i-th example. Then, given a particular aggregation operator $\mathbb{C}$ (for $\mathbb{C}$onsensus) with parameter $p \in P$, the goal is to find $p_0 \in P$ that better fits the examples. Using means square error to evaluate the fitting (the lesser the error, the larger the fitting and better the model), this problem is stated as follows: find p such that the following expression is minimized:

$$\sum_{i=1}^{M} (\mathbb{C}_p(a_1^i, \ldots, a_N^i) - b^i)^2 \qquad (4)$$

Usually, this problem has to be solved using constrained optimization meth-

ods because the parameter p has to satisfy some constraints (*e.g.* when p is a fuzzy measure monotonicity is a required condition). See *e.g.* [2] for a description of techniques for solving such problem for particular aggregation operators.

[31] considers learning one dimensional distorted probabilities through the learning of the WOWA operator [28]. This is, a distribution function **p** and a function f interpolated from a weighting vector **w** (as in Example 1) are learned. A gradient descent is applied from the optimal solution for the weighted mean and OWA. As the Choquet integral generalizes both operators, a solution for them is a feasible (usually non optimal) solution for the Choquet integral. An outline of this method is as follows:

1. p = best parameter for weighted mean

2. w = best parameter for the OWA operator

3. (p^*, w^*) = best pair from (p, w)

4. Iterate gradient descent from p^* and w^*

The best fittings for OWA and weighted mean were obtained using the methods described in [30, 32] and the gradient descent is based on [6]. See [31] for details.

Our approach for learning two dimensional distorted probabilities starts from the one dimensional distorted probability $\mu = f \circ p$. Then, for a given partition of the variables X into two disjoint sets X_1 and X_2 we define a two dimensional fuzzy measure μ_{2D} that is equal to μ for all subsets of X.

Naturally, μ_{2D} is defined in terms of two probability distributions p_1 and p_2 and a function $f_{2D} : [0, 1] \times [0, 1] \rightarrow [0, 1]$. The probability distributions p_1 and p_2 are defined as follows:

$$p_1(x_i) = \frac{p(x_i)}{\sum_{x \in X_1} p(x)} \text{ for all } x_i \in X_1$$

$$p_2(x_i) = \frac{p(x_i)}{\sum_{x \in X_2} p(x)} \text{ for all } x_i \in X_2$$

and the function f_{2D} is defined as a piecewise linear function based on a set of knots defining a grid structure. Such definition is based on the following construction:

Given S and T two finite and ordered sets of values in $[0, 1]$ so that $0 = s_1 < ... < s_m = 1$ and $0 = t_1 < ... < t_n = 1$, and given F a function defined over the pairs $(s, t) \in S \times T$, we define $f_{S,T,F}$ in $[0, 1] \times [0, 1]$ as follows:

$$f_{S,T,F}(s,t) = \beta*(\alpha*F(s_i,t_j)+(1-\alpha)*F(s_{i+1},t_j))+(1-\beta)*(\alpha*F(s_i,t_{j+1})+(1-\alpha)*F(s_{i+1},t_{j+1}))$$

where,

- $s_i \in S$ is such that $s_i \leq s \leq s_{i+1}$

- $t_j \in T$ is such that $t_j \leq t \leq t_{j+1}$

- $\alpha = (s - s_i)/(s_{i+1} - s_i)$

- $\beta = (t - t_j)/(t_{i+1} - t_i)$

Then, given the 1-dimensional distorted probability μ, and the probability distributioins p_1 and p_2 defined above, we define the distortion function f_{2D} as $f_{S,T,F}$ with S, T and F defined as follows:

$S = \{0\} \cup \{p|p \in [0, 1] \text{ and there exists } Z_1 \subseteq X_1 \text{ and } p = \sum_{z \in Z_1} p_1(z)\}$

$T = \{0\} \cup \{p|p \in [0, 1] \text{ and there exists } Z_2 \subseteq X_2 \text{ and } p = \sum_{z \in Z_2} p_2(z)\}$

$F(a,b) = f(a \cdot \sum_{x \in X_1} p(x) + b \cdot \sum_{x \in X_2} p(x))$

Once p_1, p_2 and F and, thus, f_{2D} have been defined we apply a learning method for updating them and improving the fitting. The learning method consists on a two step learning process. One step is for "improving" the probability distributions p_1 and p_2 and the other is for "improving" the function F for the points in $S \times T$. Probability distributions are updated using a gradient descent approach and F is updated using a greedy method based on a random perturbation. Both steps are detailed below:

Gradient descent for updating probability distributions: This method is based on the method described in [6] for learning OWA weighting vectors. In fact, a OWA vector is equivalent to a probability distribution because weights are positive and add to one. The approach avoids dealing with weight constraints reformulating the problem so that the constraints are no longer needed. In particular, instead of considering the learning

of the weighting vector $w = (w_1 \ldots w_N)$, they consider the learning of a vector $\mathbf{\Lambda} = (\lambda_1 \ldots \lambda_N)$ from which weights are extracted as follows:

$$w = (e^{\lambda_1}/\sum_{j=1}^{N} e^{\lambda_j} \ \ldots \ e^{\lambda_N}/\sum_{j=1}^{N} e^{\lambda_j})$$

As now any vector $\mathbf{\Lambda} \in R^N$ leads to a weighting vector (positive values adding one), gradient descent can be considered for $\mathbf{\Lambda}$ without any additional constraints on Λ. We apply this approach for both weighting vectors p_1 and p_2.

Greedy method for updating the non decreasing function: This method consists on a random perturbation of the values of F on the knots. Knots $(s, t) \in S \times T$ are randomly selected and the corresponding values $F(s, t)$ are either increased or decreased also at random (but keeping monotonicity). Changes in the function are accepted only when the error decreases.

6.1 Application

We have considered the learning of two dimensional fuzzy measures for data modeling using the Choquet integral. Due to the fact that there are no public repositories with files for learning aggregation operators or information fusion models, we have used two examples taken from the machine learning repository [23]: the "iris" and the "abalone" data files. These datafiles were already used in [30], [31] and [32]. The "iris" data file consists on 150 examples each one with four input variables and one output variable. The variable *class* was recoded because it is a categorical one. The "abalone" data file consists on 4177 examples each one with 8 input variables and 1 output one. The variable *Sex* was recoded because it is a categorical one.

For the "iris" data file the results obtained with our approach are not much rellevant. Only a small improvement has been obtained (2.64903740 was the best distance – according to Expression 4 – achieved using a one dimensional distorted probability and 2.64903728 is our best result with a two dimensional one). From our point of view, this small improvement is mainly due to the fact that two of the four variables (x_1, x_4) result in probabilities equal to zero in the

model based on 1 dimensional distorted probability and the same occurs for the 2 dimensional model.

For the "abalone" data file, the best results with a one-dimensional distorted probability had a distance of 37.4931. This is achieved when four of the variables (x_2, x_5, x_6 and x_7) had a probability equal to zero. The probabilities of the variables is given in Table 6. We have considered all partitions for variables with non-null probability and the best distance is given in Table 7. In such partitions, null variables have also been included in one of the sets. The results show that in all cases, the best 2 dimensional DP also assigns probabilities equal to zero to such null variables.

Tables 8 and 9 give the probability distributions p_1, p_2 and the function F for the best solution obtained. This corresponds to the partition defined by the sets $\{x_3, x_8\}$ and $\{x_1, x_2, x_4, x_5, x_6, x_7\}$. Its best distance was 37.1682. Table 8 corresponds to the measure before the iterative process described in this section and Table 9 corresponds to the measure after the iterative process. The results show that the probability distributions p_1 and p_2 have changed, distributing part of the probability. In particular, in the dimension $\{x_1, x_4\}$ the former has lost importance and has been adquired by the latter. Similarly, the function F has also been modified.

7   Conclusions

In this work we have studied distorted probabilities. We have studied their properties and introduced a new family of measures (m-dimensional distorted probabilities). Also, we have described a method for learning two dimensional distorted probabilities and applied this method to a modeling problem. The results obtained confirm the modeling capabilities of these measures.

Acknowledgements

referees for their comments and suggestions.

References

[1] R. J. Aumann and L. S. Shapley, *Values of Non-atomic Games*, Princeton Univ. Press, 1974.

[2] G. Beliakov, (2003), How to build aggregation operators from data, Int. J. of Intel. Systems, 18 903-923.

[3] A. Chateauneuf, "Decomposable Measures, Distorted Probabilities and Concave Capacities", *Mathematical Social Sciences* 31, 19-37, 1996.

[4] G. Choquet, "Theory of Capacities", *Ann. Inst. Fourier* 5, 131-296, 1954.

[5] W. Edwards, "Probability-Preferences in Gambling," *American Journal of Psychology* 66, 349-364, 1953.

[6] D. P. Filev and R. R. Yager, "On the issue of obtaining OWA operator weights", *Fuzzy Sets and Systems*, 94, 157-169, 1998.

[7] P.C. Fishburn, "Weak qualitative probability on finite sets, *Ann. Math. Stat.* 40,2118-2126, 1969.

[8] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition", *Proc. of the Int. Joint Conf. of the 4th IEEE Int. Conf. on Fuzzy Systems and the 2nd Int. Fuzzy Engineering Symposium*, 145–150, Yokohama, Japan, 1995.

[9] M. Grabisch, "Fuzzy integral in multicriteria decision making", *Fuzzy Sets and Systems* 69, 279-298, 1995.

[10] M. Grabisch, "k-order additive fuzzy measures", *Proc. 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, 1345-1350, Granada, Spain, 1996.

[11] M. Grabisch, "Modelling data by the Choquet integral", in V. Torra (Ed.), *Information Fusion in Data Mining*, Springer, 135-148, 2003.

[12] M. Grabisch, H. T. Nguyen and E. A. Walker, *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*, Kluwer Academic Publishers, Dordreht, The Netherlands, 1995.

[13] J. Handa, "Risk, Probabilities, and a New Theory of Cardinal Utility," *Journal of Political Economy* 85, 97-122, 1977.

[14] A. Honda, T. Nakano, Y. Okazaki, "Distortion of fuzzy measures", *Proc. of the SCIS/ISIS conference*, 2002.

[15] A. Honda, T. Nakano and Y. Okazaki, "Subjective evaluation based on distorted probability", *Proc. of the SCIS/ISIS conference*, 2002.

[16] H. Imai, M. Miyamori, M. Miyakosi and Y. Sato, "An algorithm Based on Alternative Projections for a Fuzzy Measures Identification Problem", *Proc. Int. Conf. Soft Comp.*, Iizuka, Japan, 2000.

[17] D. Kahneman and A. Tversky , "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263-291, 1979.

[18] J.M. Keller, P.D.Gader and A. K. Hocaoğlu, Fuzzy integral in image processing and recognition, in Grabisch, Michel, Toshiaki Murofushi and Michio Sugeno (Eds.), *Fuzzy Measures and Integrals: Theory and Applications."* Physica-Verlag, Berlin, 2000.

[19] C. H. Kraft, J. W. Pratt and A. Seidenberg "Intuitive probability on finite sets", *Ann. Math. Statist.* 30, 408-419, 1959.

[20] J.-L. Marichal and M. Roubens, "Determination of weights of interacting criteria from a reference set", *European Journal of Operational Research* 124:3, 641-650, 2000.

[21] P. Miranda, M. Grabisch, P. Gil, (2002), p-symmetric fuzzy measures, Int. J. of Unc., Fuzz. and Knowledge-Based Systems, 10 (Supplement) 105-123.

[22] T. Murofushi, M. Sugeno, and K. Fujimoto: Separated hierarchical decomposition of the Choquet integral, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 5, no.5 (1997) pp.563-585.

[23] P. M. Murphy and D. W. Aha, "UCI Repository machine learning databases", http://www.ics.uci.edu/~mlearn/ MLRepository.html, Irvine, CA: University of California, Department of Information and Computer Science, 1994.

[24] D. Scott, "Measurement structure and Linear Inequalities", *Journal of Mathematical Psychology*, 1, 233-247, 1964.

[25] M. Sugeno, *Theory of Fuzzy Integrals and its Applications.* (PhD Dissertation). Tokyo Institute of Technology, Tokyo, Japan, 1974.

[26] M. Sugeno and T. Murofushi, it Fuzzy Measures (in Japanese), Nikkankogyou, Tokyo, Japan, 1993.

[27] A. Tanaka and T. Murofushi, "A learning model using fuzzy measure and the Choquet integral", *Proc. of the 5th Fuzzy System Symposium*, 213-217, Kobe, Japan (in Japanese), 1989.

[28] V. Torra, "The Weighted OWA operator", *Int. J. of Intel. Syst.*, 12, 153-166, 1997.

[29] V. Torra, "On hierarchically S-decomposable fuzzy measures", *Int. J. of Intel. Syst.*, 14:9, 923-934, 1999.

[30] V. Torra, "On the learning of weights in some aggregation operators: The weighted mean and the OWA operators", *Mathware and Soft Computing*, 6, 249-265, 1999.

[31] V. Torra, "Learning Weights for Weighted OWA Operators", *Proc. of the IEEE Int. Conference on Industrial Electronics, Control and Instrumentation (IECON 2000)*, Nagoya, Japan. (CD-ROM), 2000.

[32] V. Torra, "Learning weights for the quasi-weighted means", *IEEE Trans. on Fuzzy Systems*, 10:5, 653-666, 2002.

[33] V. Torra, "On some aggregation operators for numerical information", in V. Torra (Ed.), *Information Fusion in Data Mining*, Springer, 9-26, 2003.

[34] R. R. Yager, Quantifier Guided Aggregation Using OWA operators, Int. J. of Int. Systems, 11, 49-73, 1996.

[35] R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making", *IEEE Trans. on SMC*, 18, 183-190, 1988.

[36] Z. Wang and G. Klir, *Fuzzy measure theory*, Plenum Press, New York, 1992.

$Y = \{x_1, x_2, x_3, x_4, x_5\}$	$\mu_{\mathbf{p},\mathbf{w}}$	Möbius transform
{ 0 0 0 0 0}	0.0	0.0
{ 0 0 0 0 1}	0.04296875	0.04296875
{ 0 0 0 1 0}	0.1375	0.1375
{ 0 0 0 1 1}	0.2375	0.05703125
{ 0 0 1 0 0}	0.06909722	0.06909722
{ 0 0 1 0 1}	0.1375	0.02543403
{ 0 0 1 1 0}	0.3	0.09340278
{ 0 0 1 1 1}	0.5	0.07456597
{ 0 1 0 0 0}	0.18333333	0.18333333
{ 0 1 0 0 1}	0.3	0.07369792
{ 0 1 0 1 0}	0.61666666	0.29583333
{ 0 1 0 1 1}	0.7625	-0.0278646
{ 0 1 1 0 0}	0.38333333	0.13090278
{ 0 1 1 0 1}	0.61666666	0.09123264
{ 0 1 1 1 0}	0.81666666	-0.0934027
{ 0 1 1 1 1}	0.9	-0.2537327
{ 1 0 0 0 0}	0.1	0.1
{ 1 0 0 0 1}	0.18333333	0.04036458
{ 1 0 0 1 0}	0.38333333	0.14583333
{ 1 0 0 1 1}	0.61666666	0.09296875
{ 1 0 1 0 0}	0.2375	0.06840278
{ 1 0 1 0 1}	0.38333333	0.03706597
{ 1 0 1 1 0}	0.70000000	0.08576389
{ 1 0 1 1 1}	0.81666666	-0.2537326
{ 1 1 0 0 0}	0.5	0.21666667
{ 1 1 0 0 1}	0.7	0.04296875
{ 1 1 0 1 0}	0.8625	-0.2166667
{ 1 1 0 1 1}	0.93090277	-0.2537326
{ 1 1 1 0 0}	0.7625	-0.0059028
{ 1 1 1 0 1}	0.8625	-0.2537326
{ 1 1 1 1 0}	0.95703125	-0.2537326
{ 1 1 1 1 1}	1.0	0.50746528

Table 1: Fuzzy measure $\mu_{\mathbf{p},\mathbf{w}}$ and its Möbius transform. The first column denotes the set (0: the element is not included; 1: the element is included)

$\lvert X \rvert$	Number of possible orderings for Distorted Probabilities	Number of possible orderings for Fuzzy Measures
1	1	1
2	1	1
3	2	8
4	14	70016
5	546	$\mathcal{O}(10^{12})$
6	215470	–

Table 2: Possible orderings of the sets in 2^X with respect to the measure considering $\mu(\{1\}) \leq \mu(\{2\}) \leq \ldots$, where $\mathcal{O}(10^{12})$ is an estimation.

set	$\emptyset$	$\{2\}$	$\{1\}$	$\{3\}$	$\{1,2\}$	$\{2,3\}$	$\{1,3\}$	$\{2,1,3\}$
P	0	1/5.5	2/5.5	2.5/5.5	3/5.5	3.5/5.5	4.5/5.5	5.5/5.5
$\mu = f \circ P$	0	0	1/7	2/7	3/7	4/7	7/7	7/7

Table 3: Probability and non-decreasing function f corresponding to Example 3

set	$\{2\}$	$\{3\}$	$\{4\}$	$\{2,3\}$	$\{2,4\}$	$\{1\}$	$\{1,2\}$
μ	0.1	0.12	0.13	0.2	0.23	0.24	0.25

$\{3,4\}$	$\{5\}$	$\{2,3,4\}$	$\{1,3\}$	$\{1,4\}$	$\{2,5\}$	$\{1,2,3\}$	$\{3,5\}$	$\{1,2,4\}$
0.35	0.4	0.42	0.43	0.5	0.51	0.52	0.53	0.58

$\{4,5\}$	$\{1,3,4\}$	$\{2,3,5\}$	$\{2,4,5\}$	$\{1,5\}$	$\{1,2,3,4\}$	$\{1,2,5\}$	$\{3,4,5\}$
0.60	0.61	0.62	0.63	0.65	0.70	0.73	0.76

$\{2,3,4,5\}$	$\{1,3,5\}$	$\{1,4,5\}$	$\{1,2,3,5\}$	$\{1,2,4,5\}$	$\{1,3,4,5\}$
0.80	0.82	0.83	0.85	0.90	0.92

$\mu(\emptyset) = 0$	$\mu(X) = 1$
$\mu(\{M\}) = 0.45$	$\mu(\{P, L\}) = 0.9$
$\mu(\{P\}) = 0.45$	$\mu(\{P, M\}) = 0.5$
$\mu(\{L\}) = 0.3$	$\mu(\{L, M\}) = 0.9$

Table 4: Grabisch fuzzy measure in [9]

$$
\begin{array}{ccccc}
a_1^1 & a_2^1 & \ldots & a_N^1 & | & b^1 \\
a_1^2 & a_2^2 & \ldots & a_N^2 & | & b^2 \\
\vdots & \vdots & & \vdots & & \vdots \\
a_1^M & a_2^M & \ldots & a_N^M & | & b^M
\end{array}
$$

Table 5: Data examples.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
p	0.018	0.0	0.369	0.524	0.0	0.0	0.0	0.088

Table 6: Probability distribution for the variables in the "abalone" data file.

Partition	Min. Distance
$\{x_3, x_4, x_8\}, \{x_1, x_2, x_5, x_6, x_7\}$	37.3569
$\{x_4, x_8\}, \{x_1, x_2, x_3, x_5, x_6, x_7\}$	37.2745
$\{x_3, x_8\}, \{x_1, x_2, x_4, x_5, x_6, x_7\}$	37.1682
$\{x_3, x_4\}, \{x_1, x_2, x_5, x_6, x_7, x_8\}$	37.2751
$\{x_8\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$	37.2866
$\{x_4\}, \{x_1, x_2, x_3, x_5, x_6, x_7, x_8\}$	37.2634
$\{x_3\}, \{x_1, x_2, x_4, x_5, x_6, x_7, x_8\}$	37.2810

Table 7: Partitions considered with the "Abalone" data file and minimum distance achieved

1.0	0.457	0.475	0.981	1.0
0.806	0.369	0.387	0.893	0.911
0.193	0.088	0.106	0.612	0.630
0.0	0.0	0.018	0.524	0.542
	0.0	0.033	0.966	1.0

p_1	$p_1(x_1) = 0.033$	$p_1(x_2) = 0$	$p_1(x_4) = 0.966$	$p_1(x_5) = 0$	$p_1(x_6) = 0$	$p_1(x_7) = 0$
p_2	$p_2(x_3) = 0.806$	$p_2(x_8) = 0.193$				

Table 8: Initial function F on the knots defined by $S = \{0, 0.033, 0.966\}$ and $T = \{0, 0.193, 0.806, 1.0\}$ (top) and probability distributions p_1 and p_2 (bottom) for the partition defined by $\{x_3, x_8\}$ and $\{x_1, x_2, x_4, x_5, x_6, x_7\}$

0.806	0.351	0.390	0.698	1.0
0.193	0.351	0.390	0.512	0.595
0.0	0.0	0.016	0.016	0.016
	0.0	0.033	0.966	1.0

p_1	$p_1(x_1) = 0$	$p_1(x_2) = 0$	$p_1(x_4) = 1.0$	$p_1(x_5) = 0$	$p_1(x_6) = 0$	$p_1(x_7) = 0$
p_2	$p_2(x_3) = 0.981$	$p_2(x_8) = 0.018$				

Table 9: Final function F on the knots defined by $S = \{0, 0.033, 0.966\}$ and $T = \{0, 0.193, 0.806, 1.0\}$ (top) and probability distributions p_1 and p_2 (bottom) for the partition defined by $\{x_3, x_8\}$ and $\{x_1, x_2, x_4, x_5, x_6, x_7\}$