Indistinguishability relations in Dempster–Shafer theory of evidence

Enric Hernández a,*, Jordi Recasens b

a Secció de Matemàtiques i Informàtica, ETSAB, Universitat Politècnica de Catalunya, Avda. Diagonal, 649, 08028 Barcelona, Spain
b Secció de Matemàtiques i Informàtica, ETSAV, Universitat Politècnica de Catalunya, Pere Serra 1-15, 08190 Sant Cugat, Spain

Received 1 July 2003; accepted 1 February 2004
Available online 28 April 2004

Abstract

Each theory or model implicitly defines its inherent notion of equality for the objects in question. In turn, this equality, and its counterpart, the mathematical concept of equivalence, provides the basis on which to establish classification mechanisms for the domain at hand.

Nevertheless, equivalence relations have not been proved to be sufficiently suited to capturing the underlying structure when dealing with domains pervaded with uncertainty. Therefore, the need for a more flexible definition of equality brings the concept of \( T \)-indistinguishability operator on to the scene.

In this paper, we study the notion of indistinguishability within the context of the Dempster–Shafer theory of evidence and provide effective definitions and procedures for computing the \( T \)-indistinguishability operator associated with a given body of evidence. We also show how these procedures could also be adapted in order to provide a new method for tackling the problem of belief function approximation based on the concept of \( T \)-preorder.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Dempster–Shafer theory of evidence; \( T \)-indistinguishability operator; \( T \)-preorder; Belief function approximation

*Corresponding author.

E-mail addresses: enriche@lsi.upc.es (E. Hernández), j.recasens@upc.es (J. Recasens).

0888-613X/S - see front matter © 2004 Elsevier Inc. All rights reserved.
1. Introduction

The significance of classification, a key task underlying most cognitive activities that require a high level of abstraction or generalization, such as the formulation of models and theories, is today widely recognized.

Moreover, classification as the process of grouping or clustering according to a certain criterion of similarity tends to be intimately related to traditional notions of identity, indiscernibility and indistinguishability. All these concepts have a long tradition as subjects of discussion within a great number of fields, ranging from philosophy and psychology to mathematics.

The standard way of approaching the concept of identity in mathematics is linked to a tradition that can be traced back at least to Leibniz, whose law of identity is usually written in a second-order language as

\[ x \approx y \iff \forall P : P(x) \iff P(y) \]  

where \( x \) and \( y \) denote individuals and \( P \) ranges over the set of properties.

Leibniz’s law, which is a conjunction of the principles of the Identity of Indiscernibles and Indiscernibility of Identicals, is intended to express the concept of identity as agreement with respect to all properties. The original postulate may have evolved towards more elaborated formulations based on the idea of the invariance of the set of all automorphisms definable over a given structure, but the main idea behind remains the same.

When all the properties involved are entirely precise (lack of uncertainty), what we obtain is the classic equality, where two individuals are considered equal if and only if they share the same set of properties. What happens, however, when imprecision arises, as in the case of properties which are fulfilled only up to a degree? Thus, because certain individuals will be more similar than others, the need for a gradual notion of equality arises.

A further example is when the fulfillment of properties is not a matter of degree but the limitations in perceiving and measuring these properties imply the emergence of an approximate equality. Let us, for instance, consider the case of a particular appliance providing measurements on the real line with an error margin \( \varepsilon \). It naturally defines the following approximate equality relationship:

\[ x \approx y \iff |x - y| \leq \varepsilon \]  

by which two measurements will become distinguishable only if their absolute difference is above the error threshold \( \varepsilon \).

Relation \( \approx \) is not transitive since we could have \( x \approx y, y \approx z \) and not necessarily \( x \approx z \). In [19], Poincaré was already concerned with this apparent paradox. He pointed out that equality satisfies transitivity only in the context of pure mathematics. In the real world, “equal” really means “indistinguishable”.
Lack of transitivity also appears when dealing with properties that are inherently vague, which we have previously discussed. Indeed, since a chain of objects that are usually indistinguishable can lead from one which clearly seems to be compatible with a given property to one which clearly does not, the sorites paradox and a corresponding break in transitivity must ensue.

All these considerations show that certain contexts that are pervaded with uncertainty require a more flexible concept of equality that goes beyond the rigidity of the classic concept of equality. Furthermore, since the concept of equivalence relation, as the mathematical tool that is used to define the underlying structure, is definitely lost, some workaround should be provided.

These contexts define a notion of “closeness” (accounting for the vague nature of the problem) in a natural way, which is in turn linked to some sort of underlying metrics. Thus, it seems entirely advisable to reflect this “closeness” when defining the counterpart notion of distance induced by the metric at hand.

An ideal case occurs when the distance obtained is the classic pseudo-distance \( d : X \times X \to \{0, 1\} \):

\[
\begin{align*}
(1) \quad & d(x, x) = 0, \\
(2) \quad & d(x, y) = d(y, x), \\
(3) \quad & d(x, z) \leq d(x, y) + d(y, z).
\end{align*}
\]

Then, the relation \( E \) defined by

\[
\begin{cases} 
(x, y) \in E & \text{if } d(x, y) = 0 \\
(x, y) \notin E & \text{otherwise}
\end{cases}
\]

is the classic equality (i.e. a classic equivalence relation inducing a partition on \( X \)).

Generally, however, the relation considered is not an equivalence relation. Instead, we must deal with a relation that is reflexive, symmetric and possibly not (in the usual terms) transitive. In this case, the current version of the “triangle inequality” for the underlying metrics translates into a new kind of transitivity property, in which transitivity is defined in terms of a minimum threshold.

These considerations lead to the definition of an especial type of fuzzy relation known as \( T \)-indistinguishability operator, where \( T \) is a \( t \)-norm as defined below.

**Definition 1.** A function \( T : [0, 1] \times [0, 1] \to [0, 1] \) is called a \( t \)-norm if the following conditions hold:

\[
\begin{align*}
(1) \quad & T(x, T(y, z)) = T(T(x, y), z) \quad \text{(associativity),} \\
(2) \quad & T(x, y) = T(y, x) \quad \text{(commutativity),}
\end{align*}
\]
(3) $x \leq x' \Rightarrow T(x,y) \leq T(x',y)$
$y \leq y' \Rightarrow T(x,y) \leq T(x,y')$
(monotonicity),
(4) $T(x,1) = T(1,x) = x$ (contour conditions).

**Definition 2.** A fuzzy relation $E$ on $X$ is a $T$-indistinguishability operator if and only if $\forall x, y, z \in X$ the following conditions are satisfied:

(1) $E(x,x) = 1$ (reflexivity),
(2) $E(x,y) = E(y,x)$ (symmetry),
(3) $E(x,z) \geq T(E(x,y),E(y,z))$ ($T$-transitivity).

Note that when the underlying structure induces an ultrametric, we obtain a similarity relation following Zadeh’s nomenclature [13] or a $T$-indistinguishability operator for the minimum $t$-norm. Also, reversing the restriction of Euclidean metrics into the unit interval generates a likeness relation [20] or a $T$-indistinguishability operator for the Lukasiewicz $t$-norm. Obviously, the generalized definition particularizes to the classic equality under the appropriate assumptions.

Therefore $T$-indistinguishability operator seems to be a good candidate for the more flexible and general version of the concept of equality that we are searching for.

Throughout the paper, $T$-indistinguishability operators will play a central role. One problem commonly faced when studying such relations is how to effectively build them.

The traditional approach relies on computing the transitive closure from a reflexive and symmetric relation. This method, however, has not proved to be fully satisfactory because of the computational cost involved (the closure is computed as the supremum of max-$T$ powers of the original relation) and primarily because of the distortion suffered by the initial values.

These weaknesses were surmounted by the introduction of the representation theorem for $T$-indistinguishability operators.

**Theorem 3** [6]. Let $E$ be a map from $X \times X$ into $[0,1]$ and $T$ be a continuous $t$-norm. $E$ is a $T$-indistinguishability operator if and only if a family $\{h_j\}_{j \in J}$ of fuzzy sets exists in $X$, such that

$$E(x,y) = \inf_{j \in J} \widehat{T} \left( \max(h_j(x),h_j(y)) \mid \min(h_j(x),h_j(y)) \right)$$

(3)

The preceding theorem also allows, in a natural way, the definition of the dimension of a $T$-indistinguishability operator.
**Definition 4.** A fuzzy set \( h \) on \( X \) is a generator of a \( T \)-indistinguishability operator \( E \) if \( h \) is an element of any family \( \{h_j\}_{j \in J} \) that generates \( E \) in the sense of Theorem 3.

**Definition 5.** Let \( E \) be a \( T \)-indistinguishability operator. The dimension of \( E \) is the minimum of the cardinalities of the generating families of \( E \).

In the same way, when we are concerned with the notion of order in domains pervaded with uncertainty, a natural generalization of the usual (crisp) preorder relation is given by the concept of \( T \)-preorder, which is obtained by simply “dropping” the property of symmetry from the definition of \( T \)-indistinguishability operator.

**Definition 6 [6].** A function \( P : X \times X \to [0,1] \) is a \( T \)-preorder if \( \forall x,y,z \in X \)

1. \( P(x,x) = 1 \),
2. \( T(P(x,y),P(y,z)) \leq P(x,z) \).

We also have the corresponding representation theorem and notion of dimension for \( T \)-preorder.

**Theorem 7 [6].** Let \( P \) be a map from \( X \times X \) into \( [0,1] \) and \( T \) be a continuous \( t \)-norm. \( P \) is a \( T \)-preorder if and only if a family \( \{h_j\}_{j \in J} \) of fuzzy sets exists in \( X \), such that:

\[
P(x,y) = \inf_{j \in J} \tilde{T}(h_j(x)\mid h_j(y))
\]

**Definition 8.** A fuzzy set \( h \) in \( X \) is a generator of a \( T \)-preorder \( P \) if \( h \) is an element of any family \( \{h_j\}_{j \in J} \) that generates \( P \) in the sense of Theorem 7. This theorem also gives the notion of dimension of a given \( T \)-preorder.

**Definition 9.** Let \( P \) be a \( T \)-preorder. The dimension of \( P \) is the minimum of the cardinalities of the generating families of \( P \) in the sense of the previous representation theorem.

The relative nature of the notion of equality can be inferred from the formalization of Leibniz’s law (1). Indeed, since every context or theory defines its own set of descriptive attributes (set \( P \) of properties in the aforementioned formalization), the direct application of Leibniz’s law yields different equality criteria depending on the context of discourse; therefore, every theory handles its own notion of equality over the objects in question. Biconditional operator...
and the notion of congruence as equivalence criteria for logical predicates and geometric figures respectively are just two examples.

In this paper, we aim to study the concept of indistinguishability within the framework of the Dempster–Shafer theory.

The theory of evidence constituted a generalization of the Bayesian approach to modeling subjective beliefs and overcame several of its drawbacks, such as the lack of a proper representation of ignorance or the non-symmetric property of the conditioning rule.

**Definition 10** [1]. A function $\varphi(X) \rightarrow [0, 1]$ is a belief function if and only if it satisfies the following conditions:

1. $\text{Bel}(\emptyset) = 0$,
2. $\text{Bel}(X) = 1$,
3. $\forall A_1, \ldots, A_n \subseteq X : \text{Bel}(A_1, \ldots, A_n) \geq \sum_{I \subseteq \{1, \ldots, n\}} (-1)^{|I|+1} \cdot \text{Bel}(\cap_{i \in I} A_i)$,

where $|I|$ is the cardinality of $I$.

**Definition 11** [1]. A function $m : \varphi(X) \rightarrow [0, 1]$ is called a basic probability assignment (bpa in the following) whenever

1. $m(\emptyset) = 0$ (normalization is assumed),
2. $\sum_{A \subseteq \varphi(X)} m(A) = 1$

(subsets $A \subseteq X : m(A) > 0$ are called focal elements).

**Proposition 12** [1]. Given a bpa $m$ on $X$, the associated measures of belief ($\text{Bel}$) and plausibility ($\text{Pl}$) are related univocally to each other by the following formulas:

$$\forall A \subseteq X : \text{Pl}(A) = 1 - \text{Bel}(\overline{A})$$  \hspace{1cm} (5)

$$\forall A, B \subseteq X : \text{Bel}(A) = \sum_{B | B \subseteq A} m(B)$$  \hspace{1cm} (6)

$$\forall A, B \subseteq X : m(A) = \sum_{B | B \subseteq A} (-1)^{|A - B|} \text{Bel}(B)$$  \hspace{1cm} (7)

Belief functions, as noted in [10], quantify the degree to which the evidence available supports the hypothesis that a particular element of $X$, whose characterization in terms of the relevant attributes may be deficient, belongs to a

1 In the following we will only consider domains $X$ that have a finite cardinality.
subset of $X$. This interpretation implicitly conveys the notion of indistinguishability, which is definable in terms of the compatibility between the elements of the domain and that particular element.

Moreover, from the representation of evidence by means of the definition of its basic probability mass assignment on subsets of $X$ (focal elements), it would be reasonable to explain the distinguishability between two elements $a, b$ by the exclusive support given to any of the two (defining support as holding element $a$ exclusively as a function of the masses assigned to the focal elements containing $a$ and not containing $b$, and analogously for $b$), while the inclusion of $a$ and $b$ in the same focal element should contribute to an increase in the indistinguishability between them, given that they are indistinguishable for that “portion” of evidence at least.

In this paper, we intend to explore these ideas. The paper is organized as follows: in Section 2, basic results relating to $t$-norms and the Dempster–Shafer theory of evidence are discussed; Section 3 defines $T$-indistinguishability operator $E_1$ as an approach to computing indistinguishability based on the one-point coverage function. Section 4 shows how the method followed to compute operator $E_1$ can also be used to tackle the problem of belief function approximation. A new approach based on the concept of $T$-preorder is proposed. In Section 5, a new $T$-indistinguishability operator ($E_2$) is introduced in order to overcome some of the drawbacks of the previously defined operator $E_1$. Sections 6 and 7 afford better insight into the study of operator $E_2$ by addressing topics such as dimensionality. Finally, our conclusions are presented.

2. Preliminaries

This section provides several definitions and propositions that will be used throughout the paper. Some of these results are well known but they are included in order to make this work as self-contained as possible.

2.1. On $t$-norms and generalities

**Definition 13.** A $t$-norm $T$ is Archimedean if and only if the set \( \{ x \in [0, 1] : T(x, x) = x \} \) equals \{0, 1\}.

**Theorem 14** [21]. A continuous $t$-norm $T$ is Archimedean if and only if a strictly decreasing continuous function $t : [0, 1] \rightarrow [0, +\infty]$ with $t(1) = 0$ exists, such that

$$T(x, y) = t^{-1}(t(x) + t(y))$$  \hspace{1cm} (8)
where \( t^{-1} \) is the pseudo-inverse of \( t \) defined as

\[
t^{-1}(x) = \begin{cases} 
  t^{-1}(x) & \text{if } x \in [0, t(0)] \\
  0 & \text{otherwise}
\end{cases}
\] (9)

Then function \( t \) is called an additive generator of the \( t \)-norm \( T \).

**Definition 15.** A \( t \)-norm \( T \) is strict if the set \( \text{Nil}T \) defined as

\[
\{x \in (0, 1) : \exists m \in N \text{ such that } T^m(x) = 0\}
\] (10)
equals \( \emptyset \), and non-strict if \( \text{Nil}T = (0, 1) \). \(^2\)

**Definition 16.** Given a continuous \( t \)-norm \( T \), its residuation \( \hat{T} \) is defined as:

\[
\forall x, y \in [0, 1] : \hat{T}(x|y) = \sup \{z \in [0, 1] : T(z, x) \leq y\}
\] (11)

Let us present the residuations for the three most commonly used \( t \)-norms.

1. When \( T(x, y) = \min(x, y) \) then

\[
\hat{T}(x|y) = \begin{cases} 
  1 & x \leq y \\
  y & \text{otherwise}
\end{cases}
\] (12)

2. When \( T(x, y) = x \cdot y \) then

\[
\hat{T}(x|y) = \min \left(1, \frac{y}{x}\right)
\] (13)

3. When \( T(x, y) = \max(x + y - 1, 0) \) then

\[
\hat{T}(x|y) = \min(1 - x + y, 1)
\] (14)

**Definition 17.** Given a continuous \( t \)-norm \( T \), its biresiduation \( \quad \not\hat{T} \) is defined as:

\[
\quad \not\hat{T}(x, y) = \min(\hat{T}(x|y), \hat{T}(y|x))
\] (15)

\(^2\) \( T^m \) is defined by recursion as \( T^1(x) = x \) and \( T^m(x) = T(T^{m-1}(x), x) \).
2.2. On Dempster–Shafer theory of evidence

**Theorem 18** [1]. Let $m$ be a bpa on $\varphi(X)$ and $Pl$ be its associated plausibility measure. Then $Pl$ is a possibility measure if and only if the family of focal elements of $\varphi(X)$ is nested.

**Definition 19** [7]. Every possibility measure $\text{Pos}$ on $\varphi(X)$ can be uniquely determined by a possibility distribution function $h : X \rightarrow [0, 1]$, such that

$$\forall A \in \varphi(X) : \text{Pos}(A) = \max_{x \in A} h(x) \quad (16)$$

**Definition 20.** Let $m \in M$ be a bpa on $X$. We will say that $m$ is consistent when

$$\bigcap_{A \subseteq X, m(A) > 0} A \neq \emptyset \quad (17)$$

3. A projection-based approximation to the definition of indistinguishability

In this section, we present a method to calculate the $T$-indistinguishability operator associated with a given body of evidence which stems from a particular restriction of the plausibility measure over the set of singletons.

3.1. Covering functions

The general concept of the covering function comes from “projecting” a measure defined in $\varphi(X)$ over a subset $S$ of $\varphi(X)$. We will have different types of covering functions depending on $S$ and the definition of the projection function.

In this paper, we will use the notion of the commonality number introduced by Shafer [1] to define a specific type of covering function.

**Definition 21** [1]. Let $m$ be a basic probability assignment in $\varphi(X)$. The commonality number $Q : \varphi(X) \rightarrow [0, 1]$ associated with $m$ is defined as:

$$\forall A \in \varphi(X) : Q_m(A) = \sum_{B \in \varphi(X) | A \subseteq B} m(B) \quad (18)$$

Quoting Shafer: “$Q(A)$ measures the total probability mass that can move freely to every point of $A$”. Using this definition and following Goodman [2]:

**Definition 22** [2]. Let $\varphi_n(X) = \{B : B \in \varphi(X) \wedge |B| \leq n\}$, $n \geq 1$, and $m$ a bpa in $\varphi(X)$. Then, the $n$-point coverage function of $m$ is defined as:

$$\forall A \in \varphi_n(X) : m_n(A) = Q_m(A) \quad (19)$$
By varying \(1 \leq n \leq |X|\), we obtain different projections of the measure \(Q_m\). The case \(n = 1\) has received especial attention in the literature [2–5].

**Definition 23.** Let \(m\) be a bpa in \(\wp(X)\). Its one-point coverage function \(\mu_m : X \rightarrow [0, 1]\) \(^3\) is defined as:

\[
\forall x \in X : \mu_m(x) = Q_m(\{x\}) = \sum_{B \subseteq \wp(X) : x \in B} m(B)
\]

(20)

\(\mu_m\) is the projection of \(Q_m\) over the singleton set, that is, the amount of mass that can be moved to every element \(x\) of \(X\).

It is obvious that, because \(\text{Pl}\) is the plausibility measure associated with \(m\), then:

\[
\forall x \in X : \mu_m(x) = Q_m(\{x\}) = \text{Pl}(\{x\})
\]

(21)

The equality above provides an interpretation of \(\mu_m\) as a fuzzy set, in which the membership degree represents the compatibility between a particular element and the evidence, computed as the sum of the masses of the focal elements that are compatible with the element in question.

### 3.2. T-indistinguishability operator \(E_1\)

The idea behind \(T\)-indistinguishability operator \(E_1\) is based on using the previously introduced one-point coverage function as an approximation of the original bpa in order to generate the intended indistinguishability.

**Proposition 24** [6]. *Given a fuzzy set \(\mu\) in \(X\), the fuzzy relation \(E\) defined \(\forall x, y \in X\) as*

\[
E(x, y) = T(\mu(x), \mu(y))
\]

(22)

*is a \(T\)-indistinguishability operator.*

Subsequently, we offer a formal definition of the idea sketched above.

**Definition 25.** Let \(m\) be a bpa in \(X\) and \(\mu_m\) its one-point coverage function. The \(T\)-indistinguishability operator \(E_1\) is defined as

\[
\forall x, y \in X : E_1(x, y) = T(\mu_m(x), \mu_m(y))
\]

(23)

\(^3\) Also called “consonant projection”, “falling shadow”, “contour function” and “point-trace”.  

---

154  
(\(E_1\) is a \(T\)-indistinguishability operator as a trivial consequence of Proposition 24).

**Example 26.** Let \(m\) be the bpa in \(X = \{a, b, c, d\}\) defined by

\[
\begin{align*}
    m(\{a, c\}) &= 0.3 \\
    m(\{b, c\}) &= 0.3 \\
    m(\{a, b, c\}) &= 0.3 \\
    m(\{a, b, d\}) &= 0.1
\end{align*}
\]

Its one-point coverage function \(\mu_m\), defined as

\[
\forall x \in X : \mu_m(x) = \sum_{A \subseteq X : x \in A} m(A)
\]

is

\[
\begin{align*}
    \mu_m(\{a\}) &= 0.7 \\
    \mu_m(\{b\}) &= 0.7 \\
    \mu_m(\{c\}) &= 0.9 \\
    \mu_m(\{d\}) &= 0.1
\end{align*}
\]

Finally, \(E_1\) (taking the Lukasiewicz \(t\)-norm)

<table>
<thead>
<tr>
<th>(E_1)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>(b)</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>(c)</td>
<td>0.8</td>
<td>0.8</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>(d)</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>1</td>
</tr>
</tbody>
</table>

4. **Equivalence criteria and the issue of belief function approximation**

This section is devoted to showing how the aforementioned procedure for computing the \(T\)-indistinguishability operator \(E_1\) associated with a given body of evidence may be used to introduce a new approach to the well-known problem of belief function approximation.

After a short presentation of several significant references to the aforementioned problem, we show how the concept of \(T\)-preorder (generated, as is operator \(E_1\), from the one-point coverage function of the belief function considered) is better suited than \(T\)-indistinguishability operators when dealing with the uniqueness of the procedure for computing the approximation.
4.1. Previous work

The issue of approximating a given belief measure has been addressed by several authors. The need for such approximations is due to the high computational cost of managing such measures.

Indeed, given a frame of discernment $X$, a mass function can have up to $2^{|X|} - 1$ focal elements, all of which must be represented explicitly in order to properly capture and combine the evidence they encode.

An approximation of a given belief measure is expected to be simpler and well-suited to computation and explanation concerns. A natural way of simplifying a given measure is to reduce the number of focal elements, which, roughly speaking, may be accomplished by “dropping” the less informative focal elements (for instance, those with smaller masses) or by using, as approximations, a special kind of evidence that can be expressed in terms of a distribution defined in the frame of discernment $X$. As Dubois and Prade point out [3], on a set $X$ whose cardinality is $n$, we need $2^n - 2$ values to define a belief or a plausibility measure from a bpa, while $n - 1$ (taking into account the normalization condition) values are enough to define a fuzzy measure that can be expressed in terms of a distribution.

Examples of this approach are the k-l-x method [15], summarization and the $D1$ approximation algorithm [17].

With reference to the latter approach, two obvious distribution-based measures have been suggested as candidates: possibility and probability measures. As far as probabilistic approximations are concerned, the pignistic approximation of a given basic probability assignment $m$ in the frame of discernment $X$ defined by:

$$\forall x \in X : p(x) = \sum_{A \subseteq X, x \in A} \frac{m(A)}{|A|}$$

and the one proposed by Voorbraak [15]

$$\forall x \in X : p(x) = \frac{\sum_{A \subseteq X, x \in A} m(A)}{\sum_{B \subseteq X} m(B) \cdot |B|}$$

are worthy of mention.

Consonant approximations have been studied in detail by Dubois and Prade [9]. In their paper, they provide effective procedures for computing inner and outer consonant approximations based on the concepts of weak and strong inclusion between random sets.

It should be noted that approximating a general belief measure by means of a simpler one is not free of consequences: it implies a reduction or loss of information. Therefore, the question becomes which properties to preserve,
and the answer may range from committing an approximation to preserve the amount of uncertainty, based on the principle of uncertainty invariance stated by Klir [11] to methods that preserve certain coherence principles such as “only the probable is possible” (for probability–possibility transformations [12]) and the aforementioned concept of weak and strong inclusion.

In this work, we present an approximation method based on a new concept: the preservation of the $T$-preorder defined by the compatibility degree between the evidence and the singleton set.

4.2. Tackling order and uniqueness concerns

From the previous section, we may conclude that the problem of approximating a given belief function is reduced to providing a simpler approximation whilst ensuring that certain restrictions are fulfilled. These restrictions range from limiting the number of focal elements [15–18] to more sophisticated methods, such as the upper (plausibility) and lower (belief) expectations interval inclusion requirements [9] or the fulfillment of Klir’s uncertainty invariance principle [11], among others.

Therefore, selecting an approximation method becomes, in a sense, a case of deciding which of the properties conveyed by the evidence should be maintained.

The approach followed to compute $E_1$ allows us to define a partition in the set of all bpa where each class of equivalence contains all bpa generating the same $T$-indistinguishability operator $E_1$, thereby preserving the property of being equivalent with respect to their associated $T$-indistinguishability operator when restricting evidence to the singleton set.

Therefore, given a bpa $m$, any other bpa belonging to its class of equivalence could be considered a candidate for its approximation. Since we are interested in “simple” approximations (simpler, at least, than the original bpa), we should search among its class of equivalence for “good”, simple candidates. Once again, distribution-based (possibilistic and probabilistic) representations of evidence stand as the best choice. Besides, uniqueness is encouraged in order to make the selection process deterministic.

Unfortunately, it is not true to say that classes of the quotient set (for the above equivalence relation) do have a unique possibilistic canonical element as shown by the following counter-example:

**Example 27.** Let $m$ be the following bpa:

\[
\begin{align*}
m(\{b\}) &= 0.1 \\
m(\{a, c\}) &= 0.6 \\
m(\{a\}) &= 0.3
\end{align*}
\]
Note that \( m \) is neither nested nor even consistent (as defined by Definition 20). Let \( m', m'' \) be defined as

\[
\begin{align*}
m'(\{a\}) &= 0.3 \\
m'(\{a, c\}) &= 0.5 \\
m'(\{a, b, c\}) &= 0.2 
\end{align*}
\]

and

\[
\begin{align*}
m''(\{b\}) &= 0.5 \\
m''(\{b, c\}) &= 0.3 \\
m''(\{a, b, c\}) &= 0.2 
\end{align*}
\]

Since all \( m, m', m'' \) generate the same \( T \)-indistinguishability operator \( E_1 \) (assuming the Lukasiewicz \( t \)-norm), namely

<table>
<thead>
<tr>
<th></th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>1</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>( b )</td>
<td>0.2</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>( c )</td>
<td>0.7</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

and since \( m' \) and \( m'' \) are nested (i.e. both define an associated plausibility which is a possibility measure (18)), clearly \( m, m', m'' \) belong to the same class of equivalence, both \( m' \) and \( m'' \) being possibilistic evidences related to \( m \).

As a result, the uniqueness of the possibilistic canonical representative must be discarded.

Moreover, certain areas of application require not just the relative notion of indistinguishability but the more restrictive concept of order to be preserved. For instance, dealing with a decision-making problem usually involves ranking the set of different alternatives in order to choose the best one according to a predefined criterion. In other words, we are interested in obtaining the order implicitly defined by the evidence.

All these reasons lead us to consider the notion of order as the key property to be preserved by any approximation and \( T \)-preorder as the appropriate mathematical instrument for dealing with it. The subsequent section will develop this idea further.

### 4.3. Equivalence criteria

As we have pointed out, any non-trivial approximation of a measure involves a simplification or loss of information and, at the same time, enables equivalence criteria for different bodies of evidence to be established.
In this section, we take a closer look at this idea and emphasize the fact that approximations should be informative enough to provide an order on the elements of the domain \(X\), according to their compatibility with the evidence.

The one-point coverage function seems adequate for this purpose. As previously stated, this function measures the compatibility of each element with the evidence by means of the definition of a fuzzy set; consequently, its membership function (as any other membership function) enables the definition of a natural preorder \(\leq_{\mu_m}\) between the elements of the domain. This preorder can be defined, in terms of the membership function and the usual order in the unit interval \(\leq [0,1]\), as:

\[
\forall x, y \in X : x \leq_{\mu_m} y \iff \mu_m(x) \leq [0,1] \mu_m(y)
\]

(26)

Any equivalence criterion between bpa will be required to at least preserve the preorder above between the elements of the domain, that is, any two equivalent bpa \(m, m'\) should define the same preorder \(\leq_{\mu_m} = \leq_{\mu_{m'}}\).

Observing this precept, let us consider the following three criteria.

Let \(m, m'\) be two bpa in \(\wp(X)\) and \(\mu_m\) and \(\mu_{m'}\) their one-point coverage functions respectively. Then:

- \(m\) and \(m'\) are equivalent if and only if their one-point coverage functions are equal.
- \(m\) and \(m'\) are equivalent if and only if the \(T\)-preorders generated by \(\mu_m\) and \(\mu_{m'}\) are equal.
- \(m\) and \(m'\) are equivalent if and only if the natural preorders (crisp) \(\leq_{\mu_m}\) and \(\leq_{\mu_{m'}}\) defined by \(\mu_m\) and \(\mu_{m'}\) are equal.

In order to formalize these definitions, we introduce the following lemma:

**Lemma 28** [6]. Any fuzzy set \(\mu\) on \(X\) generates a \(T\)-preorder \(P_\mu\) in \(X\) in the following form:

\[
P_\mu(x, y) = \tilde{T}(\mu(x)|\mu(y))
\]

(27)

This lemma, together with the definitions regarding the concept of \(T\)-preorder presented in the introduction, allow us to formally define the equivalence criteria that we stated previously.

**Definition 29.** Let \(M\) be the set of all bpa on \(\wp(X)\); \(m \in M, m' \in M\) be two bpa; \(\mu_m\) and \(\mu_{m'}\) be their one-point coverage functions; \((\leq_{\mu_m})\) and \((\leq_{\mu_{m'}})\) the preorders (crisp) on \(X\) defined by \(\forall x, y \in X:\)

\[
x \leq_{\mu_m} y \iff \mu_m(x) \leq [0,1] \mu_m(y)
\]

\[
x \leq_{\mu_{m'}} y \iff \mu_{m'}(x) \leq [0,1] \mu_{m'}(y)
\]
$P_{\mu_{m}}$ and $P_{\mu_{m'}}$ the (one-dimensional) $T$-preorders generated by $\mu_{m}$ and $\mu_{m'}$ respectively.

Then we define $R_{1}, R_{2}, R_{3} \subseteq M \times M$ as the following equivalence relations:

\begin{align*}
(m, m') \in R_{1} & \iff \forall x \in X : \mu_{m}(x) = \mu_{m'}(x) \quad (28) \\
(m, m') \in R_{2} & \iff \forall x, y \in X : P_{\mu_{m}}(x, y) = P_{\mu_{m'}}(x, y) \quad (29) \\
(m, m') \in R_{3} & \iff (\leq_{\mu_{m}}) = (\leq_{\mu_{m'}}) \quad (30)
\end{align*}

(It is trivial to check that $R_{1} \subset R_{2} \subset R_{3}$.)

**Example 30.** Let $m, m'$ be two bpa in $X = \{a, b, c\}$ defined by:

\[
m(\{a\}) = 0.2 \\
m(\{b, c\}) = 0.2 \\
m(\{a, b, c\}) = 0.6
\]

and

\[
m'(\{a\}) = 0.1 \\
m'(\{b\}) = 0.1 \\
m'(\{c\}) = 0.1 \\
m'(\{a, b, c\}) = 0.7
\]

Their one-point coverage functions are $\mu_{m}(a) = \mu_{m}(b) = \mu_{m}(c) = 0.8$ and $\mu_{m'}(a) = \mu_{m'}(b) = \mu_{m'}(c) = 0.8$ respectively.

Therefore, it holds that

\[
\forall x \in X : \mu_{m}(x) = \mu_{m'}(x) \Rightarrow (m, m') \in R_{1} \\
\forall x, y \in X : P_{\mu_{m}}(x, y) = P_{\mu_{m'}}(x, y) = 1 \Rightarrow (m, m') \in R_{2} \\
\forall x \in X : \mu_{m}(x) = \mu_{m}(x) \Rightarrow (\leq_{\mu_{m}}) = (\leq_{\mu_{m'}}) \Rightarrow (m, m') \in R_{3}
\]

**Example 31.** Let $m, m'$ two bpa in $X = \{a, b, c\}$ defined by:

\[
m(\{a, b\}) = 0.5 \\
m(\{a, c\}) = 0.25 \\
m(\{a, b, c\}) = 0.25
\]

and

\[
m'(\{a\}) = 0.7 \\
m'(\{b\}) = 0.2 \\
m'(\{c\}) = 0.1
\]
Their one-point coverage functions are
\[ \mu_m(a) = 1 \]
\[ \mu_m(b) = 0.75 \]
\[ \mu_m(c) = 0.5 \]
and
\[ \mu_{m'}(a) = 0.7 \]
\[ \mu_{m'}(b) = 0.2 \]
\[ \mu_{m'}(c) = 0.1 \]

Therefore,
\[ \forall x \in X : \mu_m(x) \neq \mu_{m'}(x) \Rightarrow (m, m') \notin R_1 \]
\[ P_{\mu_m} \neq P_{\mu_{m'}} \Rightarrow (m, m') \notin R_2 \]
\[ (\leq_{\mu_m}) = (\leq_{\mu_{m'}}) \Rightarrow (m, m') \in R_3 \]

**Proposition 32.** Let \( m_{\text{ign}} \) be the bpa on \( \wp(X) \) that represents total ignorance \( (m(X) = 1) \), \( m_{\text{unif}} \) be the bpa whose associated belief measure equals the measure of probability defined by the uniform probability distribution on \( X \). Then
\[ (m_{\text{ign}}, m_{\text{unif}}) \in R_2, R_3 \]  

(31)

**4.4. Canonical elements**

In this section, we will focus on the relation \( R_2 \). As \( M \) is the set of bpa on \( \wp(X) \), by fixing a \( t \)-norm \( T \) we define \( R_2 \in M \times M \) in the following manner:
\[ \forall m, m' \in M : (m, m') \in R_2 \iff \forall x, y \in X : P_{\mu_m}(x, y) = P_{\mu_{m'}}(x, y) \]  

(32)

It should be noted that this criterion is useful only in situations in which we are just interested in the order relation of a set of elements, which is given by their compatibility with the evidence.

If \( R_2 \) is an equivalence relation, each class of equivalence \( c \) of the quotient set \( M/R_2 \) will contain all bpa that are evidentially equivalent.

For example, let \( X \) be a set of suspects who may have committed a crime and \( M \) a set of bpa representing evidence of guilt or innocence. Then, the classes of the quotient set group evidences that will produce the same verdict, based on the ranking of guilty in the set of suspects.

Now, the question of whether a possibilistic canonical element exists for each equivalence class or not seems quite natural. Theorem 40 will provide an affirmative answer to this question.
Previous results are needed before proof can be established, however.

**Definition 33.** Let $P$ be a $T$-preorder on a set $X$. For any $x \in X$, the fuzzy subset $h_x$ defined by $h_x(y) = P(x, y) \forall y \in X$ is called a column of $P$.

**Lemma 34.** If $P$ is a one-dimensional $T$-preorder on a set $X$, then for particular $t$-norms (Archimedean and minimum), $P$ can be generated by one of its columns (which is clearly a normalized fuzzy set).

**Theorem 35.** Let $T$ be a continuous Archimedean $t$-norm, $t$ a generator of $T$ and $\mu$ and $\nu$ fuzzy subsets of $X$. Then, $\mu$ and $\nu$ generate the same $T$-preorder if and only if, $\forall x \in X$ the following condition holds:

$$t_m(x) = t_v(x) + k_1 \quad \text{with} \quad k_1 \geq \sup_{x \in X} \{- t_v(x)\}. \quad (33)$$

Moreover, if $T$ is non-strict, then $k_1 \leq \inf_{x \in X} \{t(0) - t_v(x)\}$.

**Proof.** $\Rightarrow$ Given $x, y \in X$, we can suppose $\mu(x) \geq \mu(y)$ (which implies $v(x) \geq v(y)$).

$$P_{\mu}(x, y) = \tilde{T}(\mu(x) | \mu(y)) = t^{-1}(t_m(y) - t_m(x))$$

$$P_{\nu}(x, y) = t^{-1}(t_v(y) - t_v(x))$$

where $t^{-1}$ is replaced by $t^{-1}$, because all the values in brackets are between 0 and $t(0)$.

If $P_{\mu} = P_{\nu}$, then

$$t_m(y) - t_m(x) = t_v(y) - t_v(x)$$

Therefore

$$t_m(x) - t_m(y) = t_v(x) - t_v(y)$$

Let us fix $y_0 \in X$. Then $t_m(x) = t_v(x) + t_m(y_0) - t_v(y_0) = t_v(x) + k_1$.

$\Leftarrow$ Trivial. $\Box$

**Corollary 36.** Let $T$ be the Lukasiewicz $t$-norm and $\mu$ and $\nu$ fuzzy subsets of $X$. Then, $\mu$ and $\nu$ generate the same $T$-preorder on $X$ if and only if $\forall x \in X$:

$$\mu(x) = v(x) + k \quad \text{with} \quad \inf_{x \in X} \{1 - v(x)\} \geq k \geq \sup_{x \in X} \{- v(x)\}. \quad (34)$$

**Proof.** With the same notations as those of the previous theorem and taking $t(x) = 1 - x$ as a generator of the $t$-norm,

$$1 - \mu(x) = 1 - v(x) + k_1 \quad \text{with} \quad \sup_{x \in X} \{- 1 + v(x)\} \leq k_1 \leq \inf_{x \in X} \{v(x)\}$$
and therefore
\[ \mu(x) = v(x) + k \quad \text{with } \inf_{x \in X} \{1 - v(x)\} \geq k \geq \sup_{x \in X} \{-v(x)\} \]

**Corollary 37.** Let \( T \) be the product \( t \)-norm and \( \mu \) and \( v \) fuzzy subsets on \( X \). Then, \( \mu \) and \( v \) generate the same \( T \)-preorder on \( X \) if and only if \( \forall x \in X \):

\[ \mu(x) = \frac{v(x)}{k} \quad \text{with } k \geq \sup_{x \in X} \{v(x)\} \quad (35) \]

**Proof.** With the same notations as those of the previous theorem and taking \( t(x) = -\ln(x) \) as a generator of the \( t \)-norm

\[ -\ln(\mu(x)) = -\ln(v(x)) + k_1 \quad \text{with } k_1 \geq \sup_{x \in X} \{\ln(v(x))\} \]

and

\[ \mu(x) = \frac{v(x)}{k} \quad \text{with } k \geq \sup_{x \in X} \{v(x)\} \]

**Proposition 38.** Let \( T \) be the minimum \( t \)-norm and let \( \mu \) be a fuzzy subset on \( X \) such that an element \( x_M \in X \) with \( \mu(x_M) \geq \mu(x) \) \( \forall x \in X \) exists. Let \( Y \subseteq X \) be the set of elements \( x \) of \( X \) with \( \mu_*(x) = \mu(x_m) \) and \( s = \sup_{x \in X - Y} \{\mu_\chi\} \). A fuzzy subset \( v \) on \( X \) generates the same \( T \)-preorder as \( \mu \) in \( X \) if and only if:

\[ \forall x \in X - Y : \mu(x) = v(x) \wedge v(y) = \{t\} \quad \text{with } s < t \leq 1 \quad \forall y \in Y \quad (36) \]

**Proof.** It follows trivially from the fact that

\[ P_\mu(x, y) = \begin{cases} \mu(y) & \text{if } \mu(x) \geq \mu(y) \\ 1 & \text{if } \mu(x) \leq \mu(y) \end{cases} \]

**Proposition 39 [9].** Let \( m \in M \) be a bpa on \( X \) and \( \mu_m \) its one-point coverage function. Then \( \mu_m \) is normalized if and only if \( m \) is consistent.

Now we can enunciate the following theorem:

**Theorem 40.** Let \( T \) be a continuous Archimedean \( t \)-norm or the minimum \( t \)-norm, and let \( M \) be the set of bpa on \( X \). Then \( \forall m \in M \) a unique \( m' \in M \) exists, such that:
(1) \( m' \) is nested.
(2) \((m, m') \in R_2\).

**Proof.** The proof has two parts. The first one proves, in a constructive manner, the existence of \( m_0 \), and the second part deals with uniqueness.

- **Existence of \( m_0 \):** let \( \mu_m \) be the one-point coverage function of \( m \). From \( \mu_m \), we generate the \( T \)-preorder \( P_{\mu_m} \) following the method shown in Lemma 28. Obviously, \( P_{\mu_m} \) is a one-dimensional \( T \)-preorder. We denote by \( h_{P_{\mu_m}} \) the fuzzy set which corresponds to a generating column of \( P_{\mu_m} \).

\( h_{P_{\mu_m}} \) is normalized (by Lemma 34) and, consequently, defines a possibility measure \( \text{Pos} \) (see Definition 19). Then we calculate \( m_0 \) as the bpa corresponding to the measure \( \text{Pos} \). Theorem 18 ensures that \( m_0 \) is nested, and it is trivial to check that its one-point coverage function \( \mu_{m_0} \) equals the possibility distribution of the fuzzy set \( h_{P_{\mu_m}} \).

Finally, due to the fact that both \( h_{P_{\mu_m}} \) and \( \mu_m \) generate the same \( T \)-preorder \( P_{\mu_m} \), we have \((m, m') \in R_2\). □

Note that when \( m \) is not consistent, \( \mu_m \) is not normalized (see Proposition 39). In this case, the normalization strategy depending on the \( t \)-norm, \( \mu_{m'} \) corresponds to the “normalized” version of \( \mu_m \). Taking the Lukasiewicz \( t \)-norm, \( \mu_{m'} \) equals the normal version of \( \mu_m \) obtained by a normalization procedure suggested by Klir and Wierman [10], which consists in incrementing, for all \( X \), the value \( \mu_m(x) \) by the amount \( 1 - \text{height}(\mu_m) \).

For product \( t \)-norm, the resulting normalization method corresponds to the very common process of dividing by the maximum membership value \( \max_{x \in X} \{\mu(x)\} \), which clearly produces a normal distribution.

Finally, when taking min \( t \)-norm, the normalization method reduces to “raise” the membership degrees of the elements that have maximum membership value \( x \in X \), such that \( \forall y \in X : \mu_m(x) \geq \mu_m(y) \) up to 1.

These considerations should be taken as theoretical justifications for choosing the appropriate normalization procedure for a given context.

- **Uniqueness of \( m' \):** due to the fact that every continuous Archimedean \( t \)-norm is isomorphic to either the Lukasiewicz \( t \)-norm or to the product \( t \)-norm, we can restrict ourselves to these two \( t \)-norm and the minimum to prove the uniqueness of \( m' \).

Let us suppose \( n \in M \) is a nested bpa, such that \( m' \neq n \) and \((m', n) \in R_2 \). Nested bpa are a particular case of consistent bpa (since nested \( \Rightarrow \) consistent). By means of Proposition 39, we know that \( \mu_{m'} \) and \( \mu_n \) are both normalized. Besides, due to \((m', n) \in R_2 \), \( \mu_{m'} \) and \( \mu_n \) generate the same \( T \)-preorder. Then
(1) Lukasiewicz $t$-norm:
If $(m', n) \in R_2$, then Theorem 36 gives
$$\forall x \in X : \mu_n(x) = \mu_{m'}(x) + z$$
with
$$0 \geq z \geq \sup_{x \in X} \{-\mu_{m'}(x)\}$$
(a) Case $z = 0$: then $\forall x \in X : \mu_n(x) = \mu_{m'}(x)$. Because $\mu_n$ and $\mu_m$ are normalized, it follows that $n = m'$.
(b) Case $z < 0$: then
$$\forall x \in X : \mu_n(x) = \mu_{m'}(x) + z \Rightarrow$$
$$\forall x \in X : \mu_n(x) < \mu_{m'}(x) \Rightarrow$$
$\mu_n$ non-normalized $\Rightarrow$
Contradiction

(2) Product $t$-norm:
If $(m', n) \in R_2$, Theorem 37 gives
$$\mu_n(x) = \frac{\mu_{m'}(x)}{k}$$
with
$$k \geq \sup_{x \in X} \{\mu_{m'}(x)\}$$
and then
$$k \geq 1$$
(a) Case $k = 1$: then $\forall x \in X : \mu_n(x) = \mu_{m'}(x)$, and because both are normalized, it follows that $n = m'$.
(b) Case $k > 1$: then
$$\forall x \in X : \mu_n(x) < \mu_{m'}(x) \Rightarrow$$
$\mu_n$ non-normalized $\Rightarrow$
Contradiction

(3) Minimum $t$-norm:
If $(m', n) \in R_2$, Theorem 38 gives
$$\forall x \in X : \mu_n(x) = \begin{cases} 
\mu_{m'}(x) & \text{if } \mu_{m'}(x) < 1 \\
t & \text{if } \mu_{m'}(x) = 1 
\end{cases}$$
with
$$\sup_{x \in X : \mu_{m'}(x) < 1} \{\mu_{m'}(x)\} < t \leq 1$$
Then:

(a) Case \( t = 1 \): Then \( \forall x \in X : \mu_n(x) = \mu_m(x) \) and, because are both normalized, it follows that \( n = m' \).

(b) Case \( t < 1 \): Then

\[
\forall x \in X : \mu_n(x) < 1 \Rightarrow \mu_n \text{ non-normalized} \Rightarrow \text{Contradiction}
\]

Therefore, the uniqueness of \( m' \) is proved. \( \square \)

This theorem shows that, for any measure of plausibility (belief), we can find one (and just one) measure of possibility (necessity) which is evidentially equivalent when restricting the impact of evidence to the singleton set. Therefore, any evidence (represented by a bpa) can be converted into possibilistic evidence, ensuring that their compatibility ordering with the singleton set remains unchanged.

The uniqueness of this possibilistic evidence allows us to take it as the canonical element of its class of equivalence.

Once we have answered the question of the existence and uniqueness of the possibilistic canonical element affirmatively, we can begin to ask questions about the probabilistic counterpart. In other words, for any bpa \( m \), we look for a bpa \( p \in M \) which only assigns mass to singletons (and consequently define a probability measure that equals the plausibility and necessity measures) that are evidentially equivalent to \( m \).

In this case, we are able to build this (unique) bpa \( p \) when taking product \( t \text{-norm} \). For the minimum and the Lukasiewicz \( t \text{-norm} \), we must impose additional conditions in order to ensure its existence.

**Theorem 41.** Let \( M \) be the set of bpa on \( \varnothing(X) \) and \( m \in M \) and \( T \) the product \( t \text{-norm} \). Then a unique \( p \in M \) exists, such that:

1. \( p \) is a probability distribution.
2. \( (m, p) \in R_2 \).

**Proof.** Let \( m \in M \) be a bpa on \( X = \{x_1, \ldots, x_n\} \). Theorem 40 says that a unique nested \( m' \in M \), such that \( (m, m') \in R_2 \), exists.

We denote the one-point coverage function of \( m' \) by \( \mu_{m'} \). Now, from \( \mu_{m'} \), we will build a bpa \( p \) which will only assign mass to singletons, and also \( (m', p) \in R_2 \). Then we will have proved the existence of the probabilistic bpa we are looking for, because as \( R_2 \) is an equivalence relation, it holds that:
\[ (m, m') \in R_2 \quad (m', p) \in R_2 \Rightarrow (m, p) \in R_2 \]

Let us see how we might build a \( p \) from \( \mu_{m'} \). We must look for a fuzzy set \( \mu_p \), such that:

\[ \forall x \in X : \mu_p = \frac{\mu_{m'}(x)}{\alpha} \quad \text{with } \alpha \geq 1 \]

and

\[ \sum_x \mu_p(x) = 1 \]

These conditions define the following restrictions:

\[
\begin{align*}
\mu_p(x_1) &= \frac{\mu_{m'}(x_1)}{\alpha} \\
\vdots \\
\mu_p(x_n) &= \frac{\mu_{m'}(x_n)}{\alpha} \\
\mu_p(x_1) + \cdots + \mu_p(x_n) &= 1
\end{align*}
\]

A unique solution exists:

\[ \alpha = \sum_x \mu_{m'}(x) \quad (37) \]

Let \( p \) be the bpa, such that:

\[
\begin{cases}
p(A) = 0 & \forall A \subseteq X : |A| > 1 \\
p(\{x\}) = \mu_p(x) & \forall x \in X
\end{cases}
\]

Clearly, the one-point coverage of \( p \) equals \( \mu_p \), and given that \( \mu_p = \frac{\mu_{m'}(x)}{\alpha} \) with \( \alpha \geq 1 \), Corollary 37 allows us to conclude that \( (p, m') \in R_2 \).

**Theorem 42.** Let \( M \) be the set of bpa on \( \mathcal{P}(X) \) and \( m \in M \) and \( T \) be the Lukasiewicz \( t \)-norm. Then a unique \( p \in M \) exists, such that:

1. \( p \) is a probability distribution,
2. \( (m, p) \in R_2 \),

if and only if

\[ \frac{\sum_x \mu_{m'}(x) - 1}{|X|} \leq \inf_{x \in X} \{\mu_{m'}(x)\} \quad (38) \]
\(\mu_{m'}\) being the one-point coverage function of the unique nested \(m' \in M\), such that \((m, m') \in R_2\).

**Proof.** Following the same reasoning used in the proof of Theorem 41, we look for a fuzzy set \(\mu_p\), such that:

\[
\forall x \in X : \mu_p = \mu_{m'}(x) + \alpha \quad \text{with} \quad 0 \geq \alpha \geq \sup_x \{-\mu_{m'}(x)\}
\]

and

\[
\sum_x \mu_p(x) = 1
\]

These two conditions define the following restrictions:

\[
\begin{align*}
\mu_p(x_1) &= \mu_{m'}(x_1) + \alpha \\
&\vdots \\
\mu_p(x_n) &= \mu_{m'}(x_n) + \alpha \\
\mu_p(x_1) + \cdots + \mu_p(x_n) &= 1
\end{align*}
\]

These restrictions have a (unique) solution if and only if:

\[
\frac{\sum_x \mu_{m'}(x) - 1}{|X|} \leq \inf_{x \in X} \{\mu_{m'}(x)\}
\]  \(39\)

and the solution is

\[
\forall x \in X : \mu_p(x) = \mu_{m'}(x) + \frac{1 - \sum_x \mu_{m'}(x)}{|X|}
\]  \(40\)

Let \(p\) be the bpa, such that:

\[
\begin{cases}
p(A) = 0 & \forall A \subseteq X : |A| > 1 \\
p(\{x\}) = \mu_p(x) & \forall x \in X
\end{cases}
\]

Given that \(\mu_p = \mu_{m'}(x) + \alpha\) with \(0 \geq \alpha \geq \sup_x \{-\mu_{m'}(x)\}\) by Corollary 36, we can conclude that \((p, m') \in R_2\). \(\Box\)

**Theorem 43.** Let \(M\) be the set of bpa on \(\wp(X)\) and \(m \in M\) and \(T\) the minimum \(t\)-norm. Then a unique \(p \in M\) exists, such that:

1. \(p\) is a probability distribution,
2. \((m, p) \in R_2\),

if and only if

\[
\frac{1 - \sum_{x: \mu_{m'}(x) < 1} \mu_{m'}(x)}{\text{Card}\{x : \mu_{m'}(x) = 1\}} > \max_{x \in X} \{\{\mu_{m'}(x) : \mu_{m'}(x) < 1\}\}
\]  \(41\)
\( m' \) being the one-point coverage function of the unique nested \( m' \in M \), such that \((m, m') \in R_2\).

**Proof.** Following the argument of Theorems 41 and 42, let us define \( \mu_p \) as the fuzzy set, such that:

\[
\forall x \in X : \mu_p(x) = \begin{cases} 
\mu_{m'}(x) & \mu_{m'}(x) < 1 \\
1 - \sum_{x : \mu_{m'}(x) < 1} \mu_{m'}(x) & \mu_{m'}(x) = 1 
\end{cases}
\]

and

\[
\sum_x \mu_p(x) = 1
\]

with

\[
\max_{x \in X : \mu_{m'}(x) < 1} (\mu_{m'}(x)) < t \leq 1
\]

These restrictions have a (unique) solution if and only if:

\[
\frac{1 - \sum_{x : \mu_{m'}(x) < 1} \mu_{m'}(x)}{\text{Card}(\{x : \mu_{m'}(x) = 1\})} > \max_x (\{\mu_{m'}(x) : \mu_{m'}(x) < 1\})
\]

and, in this case, the solution is:

\[
\forall x \in X : \mu_p(x) = \begin{cases} 
\mu_{m'}(x) & \mu_{m'}(x) < 1 \\
1 - \sum_{x : \mu_{m'}(x) < 1} \mu_{m'}(x) & \mu_{m'}(x) = 1 
\end{cases}
\]

Let \( p \) be the bpa, such that:

\[
\begin{cases} 
p(A) = 0 & \forall A \subseteq X : |A| > 1 \\
p(\{x\}) = \mu_p(x) & \forall x \in X
\end{cases}
\]

Then, the one-point coverage function of \( p \) equals \( \mu_p \) and, by Theorem 38, \((p, m') \in R_2\). \( \square \)

Theorem 42 has a nice geometric interpretation.

In \( X = \{a_1, a_2, \ldots, a_n\} \), every fuzzy subset \( \mu \) and probability distribution \( p \) can be identified with the points \((\mu(a_1), \mu(a_2), \ldots, \mu(a_n))\) and \((p(a_1), p(a_2), \ldots, p(a_n))\) of \([0, 1]^n\) respectively.

In Theorem 42, a probability distribution \( p \) exists if and only if

\[
\forall i = 1, 2, \ldots, n \\
\mu_p(a_i) > 0
\]

\[\text{(44)}\]

\(^4\) For the current theorem as for Theorems 41 and 42 the uniqueness of the solutions (when it exists) results from the fact that a set of restrictions with a unique solution is solved.
and
\[ \mu_{m'}(x_i) + \alpha \geq 0 \]  
(45)

and the sum of all these numbers is equal to 1.

The aforementioned geometric interpretations of fuzzy subsets of \( \mathbb{X} \) give

**Theorem 44.** Let \( m \) be a bpa on \( p(\mathbb{X}) \) and \( \mu_m \) the one-point coverage function of \( m \). A probabilistic distribution \( p \) on \( \mathbb{X} \) with \( (m, p) \in R_2 \) exists with respect to the Lukasiewicz \( t \)-norm, if and only if \( \mu_m \) belongs to the polytope of \([0, 1]^n\) defined by:

\[ \sum_{i \neq j} x_i + (1 - n) \cdot x_j \leq 1 \quad \forall j = 1, 2, \ldots, n \]  
(46)

Moreover, the probabilistic distributions \( p \) on \( \mathbb{X} \) lie on the hyperplane

\[ x_1 + x_2 + \cdots + x_n = 1 \]  
(47)

For \( n = 2 \), all classes of \( R_2 \) contain a probabilistic distribution \( p \).

It is also worth pointing out that the probability distribution \( \mu_p \), which we obtain (wherever possible) from the possibility distribution \( \mu_{m'} \), fulfills the well-known consistency criterion

\[ \forall x \in \mathbb{X} : \mu_p(x) \leq \mu_{m'}(x) \]  
(48)

in all cases.

**Corollary 45.** Let \( M \) be the set of bpa on \( \mathcal{P}(\mathbb{X}) \), and let \( m \in M \). Then, when taking product \( t \)-norm, the probabilistic approximation \( m \) computed in Theorem 41 equals Voorbraak’s [14] approximation of \( m \).

**Proof.** It follows easily if we rewrite Voorbraak’s Bayesian constant

\[ \sum_{B \subseteq \mathbb{X}} m(B) \cdot |B| \]  

in terms of ordered possibility distribution.  

\[ \sum_{B \subseteq \mathbb{X}} m(B) \cdot |B| = \sum_{i=1}^{n} (r_i - r_{i+1}) \cdot i = \sum_{i=1}^{n} r_i \]  
(49)

---

5 If we assume the finite universe \( \mathbb{X} = \{x_1, x_2, \ldots, x_n\} \) and let \( A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \), where \( A_i = \{x_1, x_2, \ldots, x_i\} \), be a complete sequence of nested subsets that contains all the focal elements of a possibility measure, we define the ordered possibility distribution as \( \{r_1, r_2, \ldots, r_n\} \) where \( r_i = \sum_{k=i}^{n} m(A_k) \). The nested structure implies that \( r_i \geq r_{i+1} \), that is, possibility distributions are, in this formulation, always ordered and \( r_1 = 1 \) and \( r_{n+1} = 0 \) (by convention) [10].
4.5. An example

Let $X = \{a, b, c, d\}$ and $m$ be the evidence represented by the following bpa:

- $m(\{a, b\}) = 0.5$
- $m(\{c, d\}) = 0.2$
- $m(\{a, b, c, d\}) = 0.3$

Taking product $t$-norm, let us build the possibilistic and probabilistic approximations of $m$.

The one-point coverage function $\mu_m$ from $m$ is defined by the following distribution:

- $\mu_m(a) = \mu_m(b) = 0.8$
- $\mu_m(c) = \mu_m(d) = 0.5$

which in turn generates the following one-dimensional $T$-preorder $P_{\mu_m}$

<table>
<thead>
<tr>
<th>$P_{\mu_m}$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1</td>
<td>1</td>
<td>0.625</td>
<td>0.625</td>
</tr>
<tr>
<td>$b$</td>
<td>1</td>
<td>1</td>
<td>0.625</td>
<td>0.625</td>
</tr>
<tr>
<td>$c$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$d$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Theorem 40 ensures that the possibility distribution $h_{P_{\mu_m}}$ corresponding to a generating column of $P_{\mu_m}$ defines an unique nested bpa $m'$, such that $(m, m') \in R_2$.

Then, let $h_{P_{\mu_m}}$ be the possibility distribution given by:

- $h_{P_{\mu_m}}(a) = h_{P_{\mu_m}}(b) = 1$
- $h_{P_{\mu_m}}(c) = h_{P_{\mu_m}}(d) = 0.625$

Then, we compute its associated nested bpa $m'$:

- $m'(\{a, b\}) = 0.375$
- $m'(\{a, b, c, d\}) = 0.625$

which is the unique possibilistic approximation of $m$, such that $(m, m') \in R_2$.

Theorem 41 provides us with a constructive method for computing the probabilistic approximation of $m$ from the previously computed possibilistic approximation. Namely, let $\mu_{m'}$ be the one-point coverage function of $m'$ defined by:

- $\mu_{m'}(a) = \mu_{m'}(b) = 1$
- $\mu_{m'}(c) = \mu_{m'}(d) = 0.625$
and $p$ the following probability distribution:

$$
\forall x \in X : p(x) = \frac{\mu_{m'}(x)}{\sum_{y \in X} \mu_{m'}(y)}
$$

that is

$$
p(a) = p(b) = 0.3076 \\
p(c) = p(d) = 0.1923
$$

$m_p$ being the bpa defined as

$$
m_p(\{a\}) = m_p(\{b\}) = 0.3076 \\
m_p(\{c\}) = m_p(\{d\}) = 0.1923
$$

Theorem 41 ensures that $m_p$ is the unique probabilistic approximation, such that $(m,m') \in R_2$.

5. $T$-indistinguishability operator $E_2$

Returning to the problem of defining the $T$-indistinguishability operator associated with a given bpa, we should point out some of the drawbacks of the previously defined operator $E_1$.

In spite of $E_1$ satisfying the intuitive requirements as posed in the introduction, it should be noted that $E_1$ is based on an approximation of the original evidence, namely, the one-point coverage function. Despite the fact that this approximation is the optimal consonant approximation (as shown in [9]) under the weak inclusion criterion for random sets, it has the drawback inherent to any approximation consisting in the loss of information with respect to the original evidence.

Therefore, it makes sense to look for an alternative that preserves, as far as is possible, the information conveyed by the evidence. The following results will lead us to this goal.

Lemma 46 [6]. For all continuous $t$-norm $T$ and $\forall x, y, z \in X$ it holds that

$$
\overline{T}(x, z) \geq T\left(\overline{T}(x, y), \overline{T}(y, z)\right)
$$

Theorem 47. Let $F$ be a function $\varphi(X) \rightarrow [0, 1]$. Then $\forall a, b \in X$, the relation

$$
E(a, b) = \min_{A \in \varphi(X-\{a,b\})} \overline{T}(F(\{a\} \cup A), F(\{b\} \cup A))
$$

is a $T$-indistinguishability operator.
Proof. Let us prove that $E$ is reflexive, symmetric and $T$-transitive.

(a) Reflexivity. $\forall a \in X$ it holds that

$$E(a, a) = \min_{A \in \mathcal{P}(X - \{a\})} \overrightarrow{T}(F(\{a\} \cup A), F(\{a\} \cup A)) = 1$$

(b) Symmetry. Immediate from the symmetry of the operator $\overrightarrow{T}$.

(c) $T$-transitivity. By proving that $\forall a, b, c \in X$ and $\forall Z \in \mathcal{P}(X - \{a, c\})$, it holds that

$$\overrightarrow{T}(F(\{a\} \cup Z), F(\{c\} \cup Z)) \geq T(E(a, b), E(b, c))$$

then

$$E(a, c) = \min_{Z \in \mathcal{P}(X - \{a, c\})} \overrightarrow{T}(F(\{a\} \cup Z), F(\{c\} \cup Z))$$

$$\geq T(E(a, b), E(b, c))$$

and $T$-transitivity would be proved.

Let us show that the above inequality (53) is indeed satisfied. Let $Z \in \mathcal{P}(X - \{a, c\})$. Then, we can consider two cases:

(1) $b \notin Z$. Then

$$Z \in (\mathcal{P}(X - \{a, c\} \cap \mathcal{P}(X - \{b\})) = \mathcal{P}(X - \{a, b, c\})$$

and by Lemma 46 it holds that

$$\overrightarrow{T}(F(\{a\} \cup Z), F(\{c\} \cup Z))$$

$$\geq T(\overrightarrow{T}(F(\{a\} \cup Z), F(\{b\} \cup Z)), \overrightarrow{T}(F(\{b\} \cup Z), F(\{c\} \cup Z))$$

and since

$$\mathcal{P}(X - \{a, b, c\}) \subset \mathcal{P}(X - \{a, b\})$$

$$\mathcal{P}(X - \{a, b, c\}) \subset \mathcal{P}(X - \{b, c\})$$
we have
\[
T \left( \min_{V \in \wp(X - \{a,b,c\})} \tilde{T}(F(\{a\} \cup V), F(\{b\} \cup V)), \right.
\]
\[
\left. \min_{W \in \wp(X - \{a,b,c\})} \tilde{T}(F(\{b\} \cup W), F(\{c\} \cup W)) \right)
\]
\[
\geq T \left( \min_{U \in \wp(X - \{a,b\})} \tilde{T}(F(\{a\} \cup U), F(\{b\} \cup U)), \right.
\]
\[
\left. \min_{Y \in \wp(X - \{b,c\})} \tilde{T}(F(\{b\} \cup Y), F(\{c\} \cup Y)) \right)
\]
\[
= T(E(a,b), E(b,c))
\]
(2) \( b \in Z \). Then \( Z \in \wp(X - \{a, c\}) \) and besides \( Z \not\in \wp(X - \{a, b, c\}) \) which entails
\[
\{a\} \cup (Z - \{b\}) \in \wp(X - \{b,c\})
\]
\[
\{c\} \cup (Z - \{b\}) \in \wp(X - \{a,b\})
\]

By Lemma 46 it holds that
\[
\tilde{T}(F(\{a\} \cup Z), F(\{c\} \cup Z))
\]
\[
\geq T(\tilde{T}(F(\{a\} \cup (Z - \{b\}) \cup \{b\}), F(\{a\} \cup (Z - \{b\}) \cup \{c\})), \right)
\]
\[
\tilde{T}(F(\{c\} \cup (Z - \{b\}) \cup \{a\}), F(\{c\} \cup (Z - \{b\}) \cup \{b\})))
\]
\[
(54)
\]
since trivially
\[
\{a\} \cup Z = \{a\} \cup (Z - \{b\}) \cup \{b\}
\]
\[
\{c\} \cup Z = \{c\} \cup (Z - \{b\}) \cup \{b\}
\]
\[
\{a\} \cup (Z - \{b\}) \cup \{c\} = \{c\} \cup (Z - \{b\}) \cup \{a\}
\]

Moreover, as
\[
(\{a\} \cup (Z - \{b\})) \in \wp(X - \{b,c\})
\]
and
\[
(\{c\} \cup (Z - \{b\})) \in \wp(X - \{a,b\})
\]
the expression (54) holds
\[
\geq T \left( \min_{U \in \wp(X - \{b,c\})} \tilde{T}(F(U \cup \{b\}), F(U \cup \{c\})) \right.
\]
\[
\left. \min_{Y \in \wp(X - \{a,b\})} \tilde{T}(F(Y \cup \{a\}), F(Y \cup \{b\})) \right)
\]
\[
= T(E(b,c), E(a,b)) \quad \square
\]
Corollary 48. Let Bel be a belief function on X. Then, the relation
\[
E_2(a, b) = \min_{A \subseteq \nu(X - \{a, b\})} \lambda (\text{Bel}(\{a\} \cup A), \text{Bel}(\{b\} \cup A))
\]
(55)
is a T-indistinguishability operator.

Whilst the belief assigned to two equal subsets of \(X\) is obviously the same, for \(A, B \subseteq X\), such that \(A \neq B\) (and assuming \(\text{Bel}(A) \neq \text{Bel}(B)\)), the difference of belief could be “explained” by any of the differences (elements belonging to \(A\) and not belonging to \(B\), and reciprocally) between \(A\) and \(B\).

However, if we make these two sets differ exactly in just a pair of elements, that is, \(a, b \in X\) such that \(a \in A\) and \(a \notin B\), \(b \in B\) and \(b \notin A\), and \(A - \{a\} = B - \{b\} = C\), then the existing difference of belief between \(A\) and \(B\) (when it does happen) can only be explained by the differences between element \(\{a\}\) and \(\{b\}\), since the rest of elements \((C)\) are the same.

On the basis of this idea, we have defined the indistinguishability degrees for any pair \(a, b\) of elements as the minimum of the biresiduation (for a given \(t\)-norm) between their degrees of belief when both are accompanied by the same set of elements.

Let us now consider whether the definition of the \(T\)-indistinguishability operator \(E_2\) is appropriate, depending on whether it does or does not fulfill the requirement of being more informative than \(T\)-indistinguishability operator \(E_1\).

The definition of \(E_2\) does not operate with an approximation (one-point coverage function) as \(E_1\) does, but with all the information conveyed by the original belief function. Consequently, \(E_2\) is expected to be more informative (in the sense that it affords more distinguishability) than \(E_1\). The following result confirms this supposition.

Proposition 49. Let \(m\) be a bpa in \(X\) and \(E_1\) and \(E_2\) be the \(T\)-indistinguishability operators generated by Definitions 25 and 48 respectively. Then
\[
\forall x, y \in X : E_2(x, y) \leq E_1(x, y)
\]
(56)

Example 50. Let \(m\) be the same bpa as the one in Example 26
\[
m(\{a, c\}) = 0.3
\]
\[
m(\{b, c\}) = 0.3
\]
\[
m(\{a, b, c\}) = 0.3
\]
\[
m(\{a, b, d\}) = 0.1
\]
Then the $T$-indistinguishability operator $E_2$ (taking the Lukasiewicz $t$-norm) is

$$E_2(x, y) \leq E_1(x, y).$$

6. Which fuzzy measure?

A few remarks should be made regarding the generality of Theorem 47. As it does not place any restrictions on the functions it applies to (any function $\varphi(X) \to [0, 1]$ is allowed), it admits the particularization to a huge range of functions.

Nevertheless, not all these functions will provide intuitive $T$-indistinguishability operators since these functions are expected to previously convey a proper semantics (in terms of uncertain characterization), which, in a certain way, should be transferred to their associated $T$-indistinguishability operator.

Fuzzy measures, as introduced by Sugeno [22], provide a general framework for the representation of information about uncertain variables. Formally, a fuzzy measure $\mu$ on $X$ is a mapping $\mu : \varphi(X) \to [0, 1]$, such that:

1. $\mu(X) = 1$,
2. $\mu(\emptyset) = 0$,
3. $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$.

$\mu(E)$ (where $E \in \varphi(X)$) is interpreted as a measure of the “available confidence” that the uncertain value attained by a variable $V$ lies in the subset $E$. It seems pertinent, therefore, to restrict the kind of functions accepted by Theorem 47 to the more suitable class of fuzzy measures in order to grant the resulting $T$-indistinguishability operator interpretativeness in terms of the uncertain underlying structure.

The Dempster–Shafer theory also provides a framework within which information on a variable whose value is unknown may be represented. Moreover, basic probability assignments can be viewed as a structure that provides partial information on a family of fuzzy measures that are compatible with it. Typically, only two measures from this family are considered, namely the measures of belief and plausibility.

Yager [23] provides a uniform method for characterizing a family of fuzzy measures compatible with a given bpa. Let $m$ be a bpa with focal elements $B_i$, 

<table>
<thead>
<tr>
<th>$E_2$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1</td>
<td>1</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>$b$</td>
<td>1</td>
<td>1</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>$c$</td>
<td>0.7</td>
<td>0.7</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>$d$</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>1</td>
</tr>
</tbody>
</table>
For each focal element $B_i$, let $\omega_i$ be its “allocation vector” of dimension $|B_i|$, whose component $\omega_i(j)$ satisfies the following two conditions:

$$\omega_i(j) \in [0, 1]$$  \hspace{1cm} (57)

and

$$\sum_{j=1}^{|B_i|} \omega_i(j) = 1$$  \hspace{1cm} (58)

Then, a set function $\mu$ defined by

$$\forall E \in \wp(X) : \mu(E) = \sum_{j=1}^q m(B_j) \cdot \sum_{i=1}^{|B_i \cap E|} \omega_j(i)$$  \hspace{1cm} (59)

is a fuzzy measure compatible with $m$.

As Yager notes, a few especial cases are worth pointing out. If $\forall i : \omega_i(1) = 1$, then we obtain the plausibility measure; if $\forall i : \omega_i(|B_i|) = 1$, we obtain the belief measure; and if $\omega_i(j) = \frac{1}{|B_i|}$, the resulting fuzzy measure is the one used by Smets and Kennes [24].

The considerations outlined above show that, even when we are restricted to a Dempster–Shafer structure, a whole family of compatible fuzzy measures can be defined. In addition, as previously stated, the generality of Theorem 47 trivially admits the particularization of any of these measures, as with the belief measures in Corollary 48.

Why then should we favor belief measures over any other compatible measures? Plausibility measures seem to be an obvious alternative, because they are the counterparts of belief measures and are the other most common fuzzy measures associated with a given bpa.

Nevertheless, since we wish $T$-indistinguishability operators to provide as much information as possible (the more indistinguishability afforded by the operator, the more informative it will be), the following result justifies our particularization of belief measures.

**Definition 51.** Let $\text{Pl}$ be a plausibility measure on $X$. We define the $T$-indistinguishability operator $E_3$ as

$$\forall a, b \in X : E_3(a, b) = \min_{A \subseteq \wp(X \setminus \{a, b\})} \overrightarrow{T}(\text{Pl}(\{a\} \cup A), \text{Pl}(\{b\} \cup A))$$  \hspace{1cm} (60)

($E_3$ is a $T$-indistinguishability operator as a trivial corollary of Theorem 47).

**Proposition 52.** Let $m$ be a bpa over $X$ and $E_2$ and $E_3$ the $T$-indistinguishability operators as defined in Definitions 48 and 51 respectively. Then it holds that

$$\forall a, b \in X : E_2(a, b) \leq E_3(a, b)$$  \hspace{1cm} (61)
Therefore, in the subsequent section we focus on operator $E_2$. We conclude this section with a set of results which clarify the relationships between $T$-indistinguishability operators $E_1$, $E_2$ and $E_3$.

**Proposition 53.** Let $m$ be a probabilistic (i.e. a bpa which only assigns mass to singletons) bpa over $X$. Then it holds that
\[
\forall a, b \in X : E_1(a, b) = E_2(a, b) = E_3(a, b)
\]  
(62)

The preceding proposition accounts for the probabilistic case. Let us now analyze the case of possibilistic (nested) bpa.

**Proposition 54.** Let $m$ be a possibilistic (nested) bpa on $X$ and let $l_m$ be its one-point coverage function as defined in (23). Then for all $a, b \in X$, it holds that
\[
E_2(a, b) = \tilde{T}(1 - \mu_m(a), 1 - \mu_m(b))
\]
(63)
\[
E_3(a, b) = \tilde{T}(\mu_m(a), \mu_m(b)) = E_1(a, b)
\]
(64)

This proposition shows how, in the nested case, both $E_2$ and $E_3$ operators can be defined in terms of the $T$-indistinguishability operator generated by the one-point coverage fuzzy set (or its complement in the case of $E_2$). This result naturally matches our expectations, provided that, when nested, the possibility measure linked to the bpa relates biunivocally to a normal fuzzy set (its associated possibility distribution), so that the resulting indistinguishability is expected to agree with the indistinguishability generated by this fuzzy set (more precisely, by its membership degrees).

If we take the Lukasiewicz $t$-norm, we can “refine” the previous result, although we need the following lemma before doing so:

**Lemma 55** [8]. Let $T$ be the Lukasiewicz $t$-norm and $\mu$ and $\nu$ be fuzzy sets on $X$. $\mu$ and $\nu$ generate the same $T$-indistinguishability operator if and only if $\forall x \in X$
\[
\mu(x) = \nu(x) + k \quad \text{with} \quad \inf_{x \in X} \{1 - \nu(x)\} \geq k \geq \sup_{x \in X} \{-\nu(x)\}
\]
(65)

or
\[
\mu(x) = -\nu(x) + k \quad \text{with} \quad \inf_{x \in X} \{1 + \nu(x)\} \geq k \geq \sup_{x \in X} \{\nu(x)\}
\]
(66)

**Proposition 56.** Let $T$ be the Lukasiewicz $t$-norm and $m$ be a bpa on $X$. Then $\forall a, b \in X$, it holds that
\[ E_2(a, b) = E_3(a, b) \] (67)

and by Proposition 54 if \( m \) is nested it holds that

\[
E_2(a, b) = \bar{T}(1 - \mu_m(a), 1 - \mu_m(b)) \\
= \bar{T}(\mu_m(a), \mu_m(b)) \\
= E_3(a, b) = E_1(a, b)
\]

Finally, let us consider the case of ignorance. It is well known that a disadvantage of probability theory is the lack of a proper representation of ignorance, since its usual representation on the form of uniform distribution entails the acceptance of additional and unjustified assumptions. The theory of evidence overcomes this drawback by representing ignorance as the vacuous bpa \( (m(X) = 1) \). Nevertheless, it would seem desirable that—given the fact that both representations stand when we have no evidence at all that might lead to one element being favored over another—this circumstance gives no clues on how to distinguish between them based on our beliefs. The following proposition formalizes this idea:

**Proposition 57.** Let \( m \) be the bpa given by \( m(X) = 1 \) and \( p \) the probabilistic bpa given by the uniform distribution on \( X \)

\[
\forall x \in X : p(\{x\}) = \frac{1}{|X|} \] (68)

Then their associated \( E_1, E_2 \) and \( E_3 \) operators equal the trivial \( T \)-indistinguishability operator defined by

\[
\forall x, y \in X : E(x, y) = 1 \] (69)

### 7. Addressing dimensionality

The representation Theorem 3, in addition to the simplicity of the computations it involves (compared to the transitive closure approach), also provides a useful interpretation. If the family of generators are viewed as a set of features or prototypes, the theorem states that a fuzzy relation \( E \) is a \( T \)-indistinguishability operator if a set of features (whose meaning is formally defined as fuzzy sets on \( X \)) that “explains” the distinguishability between the elements in terms of their discrepancy when matching these features exists. Conversely, from a set of features we can obtain a \( T \)-indistinguishability operator that accounts for the degree of indistinguishability between the elements when only these features are taken into account.
Therefore, if we define the dimension as the minimum of the cardinalities of the generating families, it makes sense to study low dimension $T$-indistinguishability operators since these would allow the necessary computations to be simplified and, more importantly, would afford more clarity to the structure of the operator itself, because less features or prototypes would be needed to account for its indistinguishability degrees.

The simplest case occurs when the $T$-indistinguishability operator can be generated by a single feature (fuzzy set) that conveys all the information needed in such a way that, given any pair of elements, their indistinguishability degrees are defined in terms of their relative compatibility with the generating feature.

A complete set of results of the characterization of one-dimensional $T$-indistinguishability operators and effective procedures for computing the dimension and minimal families of generators of a given $T$-indistinguishability operator can be found in [25].

The purpose of this section is to perform a similar study for the $T$-indistinguishability operator $E_2$ and to provide the necessary and sufficient conditions that a given bpa must satisfy in order to generate a one-dimensional $E_2$.

### 7.1. On one-dimensional $E_2$ $T$-indistinguishability operators

Belief functions are complex mappings and are generally difficult to approximate using simpler and more understandable structures without a significant loss of information.

Nevertheless, bpa whose associated $E_2$ operator is one-dimensional can be approximated using a single feature that carries exactly the same information, from the point of view of indistinguishability, and summarizes its contents in the form of a mathematical object (fuzzy set) which allows the underlying meaning to be grasped in a more straightforward way. In other words, this fuzzy set may be considered to be the prototype that our distribution of belief is committed to.

Having discussed the motivation behind the study of one-dimensional $E_2$ operators, the first question that should be elucidated is whether exists bpa generating $E_2$ of more (than one) dimension or not, to prevent their characterization becoming trivial. In the case of $E_1$ $T$-indistinguishability operators, this characterization makes no sense since all $E_1$ are one-dimensional by definition (they are generated from one-point coverage functions and as a consequence have this function (fuzzy set) as a generator).

---

6 Despite the generality of Theorem 47, which allows a broad range of $T$-indistinguishability operators to be defined on the basis of the $t$-norm and fuzzy measure involved, from now on we will focus on the $T$-indistinguishability operator $E_2$, assuming the particularization to belief functions and the use of the Lukasiewicz $t$-norm.
The following example proves the existence of $E_2$ operators whose dimensions are greater than one.

**Example 58.** Let $m$ be the following bpa

\[
m(\{a, b, d\}) = 0.2 \\
m(\{b, c, d\}) = 0.4 \\
m(\{c, d\}) = 0.4
\]

Its associated $E_2$ operator is

<table>
<thead>
<tr>
<th>$E_2$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>$b$</td>
<td>0.6</td>
<td>1</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>$c$</td>
<td>0.4</td>
<td>0.6</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>$d$</td>
<td>0.2</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
</tr>
</tbody>
</table>

which is not one-dimensional [26].

Since the fact that nested bpa generate one-dimensional $E_2$ operators is a trivial corollary of the Proposition 54, a first attempt might involve characterizing one-dimensional $E_2$ as a certain class of bpa that satisfy well-known conditions such as nesting or consistency. The following, and previous examples, will help us to discard such an approach.

**Example 59.** Let $m$ be the bpa defined as

\[
m(\{a, b\}) = 0.2 \\
m(\{c, d\}) = 0.3 \\
m(\{d\}) = 0.5
\]

It generates the following one-dimensional $E_2$ operator

<table>
<thead>
<tr>
<th>$E_2$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1</td>
<td>1</td>
<td>0.7</td>
<td>0.2</td>
</tr>
<tr>
<td>$b$</td>
<td>1</td>
<td>1</td>
<td>0.7</td>
<td>0.2</td>
</tr>
<tr>
<td>$c$</td>
<td>0.7</td>
<td>0.7</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>$d$</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

This example proves the existence of one-dimensional $E_2$ operators whose originating bpa is neither nested nor even consistent. This, together with the
existence of consistent bpa that do not generate one-dimensional $E_2$ operators, as shown by Example 58, refutes the possibility of establishing nesting or consistency as necessary, and sufficient or necessary conditions respectively, in order to ensure one-dimensionality.

Despite our best efforts, tackling the raw problem of one-dimensional characterization directly has not proved fruitful, since it does not seem a trivial issue. A more manageable approximation that might circumvent this difficulty may involve restricting the problem to the one-dimensional characterization of certain, well-defined configurations, thereby introducing the concept of essentially one-dimensional configurations (instead of specific bpa) that are defined as subsets of the power set of $X$.

**Definition 60.** Let $F$ be a subset of the power set of $X$. We consider $F$ to be essentially one-dimensional if and only if $E_2$ is one-dimensional for all mass assignments that have $F$ as the set of focal sets.

**Example 61.** Let $X$ be a set of cardinal greater than 3 and $a, b, c \in X$. The set $F = \{\{a\}, \{c\}, \{b, c\}\}$ is not essentially one-dimensional. Consider, for example, the mass assignment

$$m(\{a\}) = 0.3$$

$$m(\{c\}) = 0.5$$

$$m(\{b, c\}) = 0.2$$

which generates the following non-one-dimensional $E_2$ operator

<table>
<thead>
<tr>
<th>$E_2$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>$b$</td>
<td>0.7</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>$c$</td>
<td>0.6</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

Nevertheless, configurations that are not essentially one-dimensional can generate a one-dimensional $E_2$ for particular mass assignments. For instance, in Example 61, consider the mass assignment

$$m(\{a\}) = 0.5$$

$$m(\{c\}) = 0.3$$

$$m(\{b, c\}) = 0.2$$
which generates the following one-dimensional $E_2$ operator

<table>
<thead>
<tr>
<th>$E_2$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>$b$</td>
<td>0.5</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>$c$</td>
<td>0.8</td>
<td>0.7</td>
<td>1</td>
</tr>
</tbody>
</table>

**Example 62.** Let $X$ be a set of cardinal greater than 3 and $a, b, c \in X$. The set $F = \{\{c\}, \{c, b\}, \{b, a\}\}$ is not essentially one-dimensional. Consider, for example, the mass assignment

\[
m(\{c\}) = 0.4 \\
m(\{c, b\}) = 0.4 \\
m(\{b, a\}) = 0.2
\]

which generates the following non-one-dimensional $E_2$ $T$-indistinguishability operator

<table>
<thead>
<tr>
<th>$E_2$</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>$b$</td>
<td>0.6</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>$c$</td>
<td>0.4</td>
<td>0.6</td>
<td>1</td>
</tr>
</tbody>
</table>

**Lemma 63.** Let $a, b \in X$ belong to exactly the same focal sets. Then $E_2(a, b) = 1$.

The next lemma proves that, as expected, nested configurations are essentially one-dimensional. Beforehand, however, we require a lemma regarding one-dimensional $T$-indistinguishability operators characterization.

**Lemma 64** [26]. A $T$-indistinguishability operator $E$ is generated by a single function $h$ if and only if there is a total order in $X(\leq)$ whose first element is $a$ and whose last element is $b$, such that for any $x, y, z \in X$ with $a \leq x \leq y \leq z \leq b$

\[
T(E(x, y), E(y, z)) = E(x, z) > 0 \tag{70}
\]

**Lemma 65.** If $F$ is nested, then $F$ is essentially one-dimensional.

**Proof.** Let $A_1 \subset A_2 \subset \cdots \subset A_s$ be the focal sets and $m(A_1), m(A_2), \ldots, m(A_s)$ their respective masses.

Let $x \in A_i - A_{i-1}$, $y \in A_j - A_{j-1}$, $z \in A_k - A_{k-1}$ with $i \leq j \leq k$. 

\[ E_2(x, y) = 1 - \sum_{l=i}^{j-1} m(A_l) \]
\[ E_2(y, z) = 1 - \sum_{l=j}^{k-1} m(A_l) \]
\[ E_2(x, z) = 1 - \sum_{l=i}^{k-1} m(A_l) \]

Therefore, for the Lukasiewicz \( t \)-norm (\( L \))
\[ L(E_2(x, y), E_2(y, z)) = E_2(x, z) \] \( \Box \)

**Lemma 66.** Let \( F = \{A_1, \ldots, A_s\} \) with \( A_i \cap A_j = A_k \cap A_l \) for all \( i, j, k, l \) with \( i \neq j, k \neq l \). Then \( F \) is one-dimensional.

**Proof.** Let \( B \) be the common intersection of the elements of \( F \). If \( x \in A_i - B \) and \( x_j \in A_j - B \), then \( E_2(x, x_j) = 1 - |m(A_i) - m(A_j)| \).

If \( x \in A_i - B \) and \( x \in B \), then \( E_2(x, x) = 1 - \sum_{j \neq i} m(A_j) \).

Let us define the following partial order in \( X \):

If \( y \in B \) then \( y \succeq x \forall x \in X \).

If \( x \in A_i - B \), \( y \in A_j - B \), then \( x \preceq y \) if and only if \( m(A_i) \leq m(A_j) \).

If \( x \not\in A_i \forall i \) then \( y \succeq x \forall y \in X \).

Therefore, if \( x \preceq y \preceq z \), then \( L(E_2(x, y), E_2(y, z)) = E_2(x, z) \). \( \Box \)

NB: If the intersection \( B \) is the empty set, then we are in the probabilistic case.

**Lemma 67.** Let \( F = \{A_1, \ldots, A_s, B\} \) with \( A_i \cap A_j = B \) for all \( i, j \) with \( i \neq j \). Then \( F \) is one-dimensional.

**Proof.** Similar to Lemma 66. \( \Box \)

**Lemma 68.** Let \( F = \{A_1, \ldots, A_s\} \) with complementary sets of \( F \) that satisfy the condition of Lemma 66. Then \( F \) is one-dimensional.

**Lemma 69.** Let \( F = \{A_1, \ldots, A_s, B\} \) with the complementary sets of \( F \) that satisfy the condition of Lemma 67. Then \( F \) is one-dimensional.

**Theorem 70.** Let \( F \) be a subset of the power set of \( X \). \( F \) is essentially one-dimensional if and only if \( F \) can be split into \( F_1, F_2, \ldots, F_s \), the sets of \( F_i \) are either nested or satisfy the conditions of one of the Lemmas 66–69 and the sets of \( F_i \) are contained in the sets of \( F_{i-1} \forall i = 2, \ldots, s \).
Proof

(\(\Leftarrow\)) Lemmas 66–69.

(\(\Rightarrow\)) (Contrareciprocal) If \(F\) cannot be split in the way required by the theorem, then either

(a) \(\exists a, c, b \in X\) with \(a \in A\), \(c \in B\), \(b \in C\) with \(A, B, C \subseteq F\) and \(a \notin B \cup C\), \(c \notin C - A\) and \(b \notin A \cup B\) or

(b) \(\exists c, b, a \in X\) with \(c \in A\), \(b \in B\), \(a \in C\) with \(A, B, C \subseteq F\) and \(c \notin C\), \(b \notin A\) and \(a \notin A\).

In case (a), let \(|F|\) denote the cardinality of \(F\). If \(|F| = 3\), then Example 62 shows that \(F\) is not essentially one-dimensional.

If \(|F| \geq 4\), let us consider the following mass assignment: \(m(A) = 0.3\), \(m(B) = 0.39\), \(m(C) = 0.3\) and for any other set \(D\) of \(F\), \(m(C) = \frac{0.01}{|F| - 3}\).

Then

\[
0.6 \leq E_2(c, b) \leq 0.61 \\
0.69 \leq E_2(c, a) \leq 0.7 \\
0.6 \leq E_2(b, a) \leq 0.61
\]

and therefore \(E_2\) is not one-dimensional.

Case (b) can be studied in a similar way. \(\Box\)

Corollary 71. \(F\) is essentially one-dimensional if and only if we cannot find cases of (a) or (b) in \(F\).

8. Conclusions

The Dempster–Shafer theory of evidence, as a framework for representing information on the degree to which available evidence is compatible with a particular element whose characterization in terms of the relevant attributes is deficient, implicitly defines a notion of indistinguishability in terms of the relative compatibility degrees between the elements of the domain and that particular element.

In this paper, we have mainly been concerned with providing definitions for the \(T\)-indistinguishability operator associated with a given body of evidence. Therefore, based on the \(T\)-indistinguishability operator generated by the one-point coverage function of the original bpa, we defined the \(T\)-indistinguishability operator \(E_1\).

In spite of operator \(E_1\) satisfying some intuitive requirements, an inherent drawback in any approximation-based approach is that it implies the loss of information with respect to the original evidence. Hence, we introduced a
general theorem providing, in a constructive manner, the $T$-indistinguishability operator associated with any function: $F : \varphi(X) \rightarrow [0, 1]$, which preserves as much of the information content as possible.

The generality of the aforementioned result allows the specialization of a huge range of functions, particularly belief functions. In this case, the resulting $T$-indistinguishability operator ($E_2$) could be considered the natural $T$-indistinguishability operator, which provides the underlying indistinguishability relation to which the distribution of belief is committed.

The characterization of one-dimensional $E_2$ operators has also been addressed, since this class of operators, in addition to affording greater clarity to the structure of the operator itself and significantly reducing the cost of computation, also enables their approximation by a single feature (generator) that carries exactly the same information from the point of view of indistinguishability as that conveyed by the original evidence.

We have also shown how the procedure for computing operator $E_1$ can be adapted in order to present a new approach to the problem of belief function approximation based on the concept of $T$-preorder.

Handling and combining belief functions involves computations that are expensive both in terms of cost and storage. It thus makes sense to provide “simpler” approximations that are better suited to computation and explanation. Our approach allowed us to group different evidences on the basis of whether they were equivalent or not when considering the impact of evidence on the definition of predefined (pre)order relations over the set of singletons. We have also provided results regarding the existence and uniqueness of possibilistic and probabilistic canonical elements for these classes of equivalence.

References