Conditional Relative Frequency Distributions with Undefined Observations and Generalized Fuzzy Orthopartitions

STEFANIA BOFFA*, IULM University, Italy
DAVIDE CIUCCI, University of Milano-Bicocca, Italy

Conditional relative frequency distributions are tools extensively employed in statistics and machine learning for analyzing connections of two or more categorical variables, examining patterns, and comparing data. As a first goal, we introduce the so-called conditional relative frequency distributions with undefined observations for representing frequencies characterized by uncertainty. After that, we show that conditional relative frequency distributions with undefined observations can be identified with particular generalized fuzzy orthopartitions, which are mathematical models describing vague partitions where the membership of elements to classes is only partially known.

JAIR Associate Editor: Prof. Chang-Dong Wang

JAIR Reference Format:
Stefania Boffa and Davide Ciucci. 2025. Conditional Relative Frequency Distributions with Undefined Observations and Generalized Fuzzy Orthopartitions. *Journal of Artificial Intelligence Research* 83, Article 13 (July 2025), 33 pages. DOI: 10.1613/jair.1.18020

1 Introduction

Conditional relative frequency distributions are tools widely spread in statistics and machine learning for exploring the relationship between two or more categorical variables, analyzing trends, and making comparisons among data [1, 2, 3, 4, 5]. This article addresses the concept of conditional frequency distributions in the presence of incomplete information, which frequently occurs in many real-world situations, such as *missing data* (some observations or variables are missing, like survey responses and sensor data), *censored data* (observations are only partially recorded, for example, the time until failure of a machine is only known if it failed during the observation period), *sampling bias* (the sample may not be representative of the full population, leading to incomplete knowledge about the underlying distribution). Several statistical techniques can be employed to estimate frequencies when data are incomplete or partially missing, including *imputation* [6, 7], *maximum likelihood estimation* [8], *Bayesian methods* [9], and *bootstrapping* [10]. Moreover, conditional relative frequency distributions practically derive from *contingency tables*, (also called *cross-tabulations* or *crosstabs*), which have been generalized to handle uncertainty in several ways; in this regard, a few examples of strategies are represented by *Probabilistic Contingency Tables* [11], *Fuzzy Contingency Tables* [12], and *Interval-based Contingency Tables* [13, 14, 15, 16].

Generalized fuzzy orthopartitions are mathematical structures classifying objects when incomplete, ambiguous, or evolving information arise [17, 18]. These extend *Ruspini partitions* [19] with partial knowledge and *orthopartitions* based on classical sets [20] with vagueness. A generalized fuzzy orthopartition is formally defined as a

*Corresponding Author.

Authors' Contact Information: Stefania Boffa, ORCID: https://orcid.org/0000-0002-4171-3459, stefania.boffa@iulm.it, IULM University, Milan, Italy; Davide Ciucci, ORCID: https://orcid.org/0000-0002-8083-7809, davide.ciucci@unimib.it, University of Milano-Bicocca, Milan, Italy.

DOI: 10.1613/jair.1.18020

particular family of intuitionistic fuzzy sets. An intuitionistic fuzzy set A is mathematically defined as a pair of functions (μ, ν) from an initial universe to [0,1] representing the degree of membership and non-membership of the objects to A. Thus, in a generalized fuzzy orthopartition, (μ, ν) represents an equivalence class C characterized by both vagueness and uncertainty, by considering that an object u of the universe belongs to C with a truth degree between $\mu(u)$ and $1 - \nu(u)$. Certainly, (μ, ν) is equivalent to a fuzzy set [21] when the degree of each element u is unambiguously determined, namely $\mu(u) = 1 - \nu(u)$. The intuitionistic fuzzy sets of a generalized fuzzy orthopartition satisfy a pair of properties capturing that the equivalent classes must be mutually disjoint and cover the initial universe. Current research has explored orthopartitions and generalized fuzzy orthopartitions in knowledge representation: in [22], entropy measures and operations are introduced for fuzzy orthopartitions; in [23], a connection between orthopartitions and possibility theory is established; [24] identifies a class of orthopartitions as special *partially-defined equivalence relations* that are equivalence relations incorporating uncertainty; [17] unifies generalized fuzzy orthopartitions and *credal partitions* that are models for partitions with partial knowledge widely used in evidential clustering [25]; in [26], operations and ordering relations are defined on generalized fuzzy orthopartitions. Moreover, generalized fuzzy orthopartitions can be understood as special *interval-valued fuzzy soft sets*, which are combinations of interval-valued fuzzy sets and soft sets [27].

As a first contribution, this article introduces novel models to describe conditional relative frequency distributions in case undefined observations are added to the initial data. More precisely, we consider a relative frequency distribution of a categorical variable Y given another one X and we suppose that a certain number of observations appear for each modality of X such that their distribution among the modalities of Y is unknown. An illustrative example is provided at the end of Section 3, where a conditional relative frequency distribution is derived from real data, which concerns the number of tourist visits to three types of attractions in each Italian region in 2015: "*museum, non-profit gallery and/or collection*", "*archaeological area or park*", and "*monument or monumental complex*". In this case, the uncertainty is introduced by assuming that 50000 new tickets are sold in every region and each of them allows a visit to only one attraction, which cannot be predicted. The structures representing conditional relative frequency distributions characterized by such type of incomplete information are called *conditional relative frequency distributions with undefined observations* and are represented as tables where cell values contain parameters subject to particular constraints. Each of these models represents a family of conditional relative frequency distributions that it coincides with, once the uncertainty is solved (namely, the distribution of the undefined observations becomes known).

Subsequently, we show that conditional relative frequency distributions with undefined observations are special cases of generalized fuzzy orthopartitions. Certainly, it is required to attach a novel semantics to generalized fuzzy orthopartitions: this time their intuitionistic fuzzy sets represent special sets of relative frequency distributions. Then, our results demonstrate that generalized fuzzy orthopartitions, originally introduced within the framework of fuzzy logic, are well-suited for describing scenarios involving frequencies rather than truth degrees. Moreover, when understanding a conditional relative frequency distribution with undefined observations as a generalized fuzzy orthopartition, an entropy measure can be introduced for quantifying its uncertainty.

Finally, we provide a sufficient and necessary condition to identify a generalized fuzzy orthopartition with a conditional frequency distribution with undefined observations. Their relationship is schematized in the Euler-Venn diagram of Figure 1. Therefore, a strength of this work is bridging fuzzy set theory with descriptive statistics through generalized fuzzy orthopartitions. Another advantage is more practical: when dealing with large sample sizes, conditional relative frequency distributions with undefined observations being represented by tables can become unwieldy and hard to interpret, and so, generalized fuzzy orthopartitions could reveal alternative, equivalent, and more convenient models. However, a first connection between orthopartitions (based on classical sets) and descriptive statistics can already be glimpsed in the literature: orthopartitions can be understood as special collections of rough sets (orthopartitions are made of orthopairs that are equivalent to rough sets), which in [28] are represented as contingency tables, although uncertainty is not factored in.

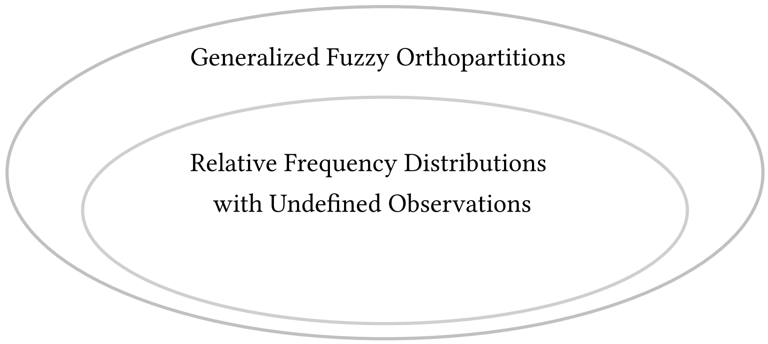

Fig. 1. Hierarchy of Generalized Fuzzy Orthopartitions and Conditional Relative Frequency Distributions with Undefined Observations.

According to the interpretation adopted in this article, generalized fuzzy orthopartitions can be viewed as tables of intervals (see Table 11), and thus understood as an extension of existing models for contingency tables with uncertainty. In particular, they generalize Probabilistic Contingency Tables [11] and Fuzzy Contingency Tables [12], by allowing the cells to contain subintervals of [0,1] rather than single-valued entries. Furthermore, although Interval-based Contingency Tables also represent cell values as intervals, the interpretation is fundamentally different: in generalized fuzzy orthopartitions, intervals reflect the uncertainty inherent in the data, whereas in interval-based scaling models, such as those found in symbolic data analysis, intervals are used to represent the internal variability of category scale values, not uncertainty [13, 29].

The article is organized as follows. The next section reviews some basic notions regarding contingency tables, relative frequency distributions (Subsection 2.1), and generalized fuzzy orthopartitions (Subsection 2.2). Section 3 has the following focal goals and its main results are schematized in Figure 4. Firstly, the notion of conditional relative frequency distributions with undefined observations is defined. Then, a particular generalized fuzzy orthopartition is constructed starting from an initial conditional relative frequency distribution with undefined observations. In addition, it is demonstrated that generalized fuzzy orthopartitions and conditional relative frequency distributions with undefined observations are equivalent models, both representing the concept of relative frequencies characterized by uncertainty. Then, important properties of generalized fuzzy orthopartitions, which are equivalent to conditional relative frequency distributions with undefined observations, are presented. Lastly, a generalized fuzzy orthopartition, equivalent to a conditional relative frequency distribution with undefined observations, is built from a real dataset. In Section 4, the inverse problem is addressed: given a fuzzy orthopartition satisfying a specific condition, the family of its equivalent conditional relative frequency distributions with undefined observations is determined. This leads to identifying a special sub-class of generalized fuzzy orthopartitions with relative frequency distributions with undefined observations. The constructions and results of Section 4 are summarized in Figures 1 and 7. Moreover, the links between the principal findings of Sections 3 and 4 are highlighted in Figure 8. Section 4 ends with a short subsection presenting four realistic application scenarios from different domains, where data uncertainty is common and generalized fuzzy orthopartitions become useful models. Finally, the last section provides some conclusions and a discuss on the possible evolutions of our results.

According to the commonly used notation, we use the symbols $\mathbb{Q}$, $\mathbb{Z}^+$, and $\mathbb{Z}_0^+$ to indicate the sets of rational numbers, positive and non-negative integers. Furthermore, all definitions and results in this article assume a finite initial universe.

We lastly provide Table 1 listing the most important symbols used in this paper (first column), the references to their definition (second column), and their brief explanation (third column).

Table 1. List of Notations.

Notation	*Definition*	*Explanation*
$\mathcal{T}$	Table 2	Contingency table of X and Y
$\mathcal{T}_{Y/X}$	Table 3	Relative frequency distribution of Y given X
t_{x_i}	Eq. (1)	Total number of observations for the modality x_i
t_{y_j}	Eq. (1)	Total number of observations for the modality y_j
(μ_A, ν_A)	Definition 2.2	Intuitionistic fuzzy set
$h_A(u)$	Eq. (2)	Degree of uncertainty
O	Definition 2.3	Generalized fuzzy orthopartition
α_i	Eq. (4)	Number of undefined observations for the modality x_i
f'_{ij}	Eq. (6)	Relative frequency of y_j given x_i with α_i undefined observations
$\mathcal{T}'_{Y/X}$	Table 7	Conditional relative frequency distribution with $\alpha_1, \ldots, \alpha_n$ undefined observations
$\Pi_{Y/X}$	Eq. (7)	The set of all conditional relative frequency distributions associated with $\mathcal{T}'_{Y/X}$
$O_{Y/X}$	Theorem 3.10	Generalized fuzzy orthopartition associated with $\mathcal{T}'_{Y/X}$
Π_O	Definition 3.12	The set of all conditional relative frequency distributions associated with a generalized fuzzy orthopartition O
$\mathcal{T}_O$	Table 18	Contingency table of X and Y associated with a generalized fuzzy orthopartition O
$\mathcal{F}_O$	Eq. (16)	Family of contingency tables of X and Y associated with a generalized fuzzy orthopartition O
$\mathcal{T}_O(a_1, \ldots, a_n)$	Table 19	Contingency table of X and Y obtained by multiplying the elements of the i-th row of $\mathcal{T}_O$ by a_i
$\mathcal{T}'_{Y/X}(a_1, \ldots, a_n)$	Definition 4.4	The relative frequency distribution of Y given X with $\alpha_1, \ldots, \alpha_n$ undefined observations deriving from $\mathcal{T}_O(a_1, \ldots, a_n)$

2 Preliminaries

Subsection 2.1 recalls some basic notions of bivariate descriptive statistics [30] and Subsection 2.2 shows the main concepts of generalized fuzzy orthopartitions that are found in [22].

2.1 Contingency Tables

We deal with bivariate data concerning two categorical variables X and Y, which respectively have the modalities $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$. Then, we write $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$.

A *contingency table* $\mathcal{T}$ is a rectangular table having n rows for the modalities of X and m columns for the modalities of Y, and its cells display the $n \times m$ possible combinations of outcomes (see Table 2). Thus, the central body of $\mathcal{T}$ is a matrix of order $n \times m$ so that its generic element a_{ij} represents the frequency of the combination of x_i and y_j, i.e. the number of times that x_i and y_i jointly occur.

Table 2. Contingency Table of X and Y.

$\mathcal{T}$	y_1	$\ldots$	y_j	$\ldots$	y_m
x_1	a_{11}	$\ldots$	a_{1j}	$\ldots$	a_{1m}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	a_{i1}	$\ldots$	a_{ij}	$\ldots$	a_{im}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_n	a_{n1}	$\ldots$	a_{nj}	$\ldots$	a_{nm}

In this article, we use the symbols t_{x_i} and t_{y_j} to denote the total number of observations for x_i and y_j, respectively. In symbols, we put

$$t_{x_i} = a_{i1} + \ldots a_{im} \quad \text{and} \quad t_{y_j} = a_{1j} + \ldots a_{nj}. \tag{1}$$

Starting from a contingency table $\mathcal{T}$, we can easily obtain the table $\mathcal{T}_{Y/X}$ (see Table 3) called *relative frequency distribution of* Y *given* X. Its i-th row displays the relative frequency distribution of Y, once each modality x_i of X is fixed. Therefore, let $j \in \{1, \ldots, m\}$, the cell of $\mathcal{T}_{Y/X}$ corresponding to x_i and y_j contains the *relative frequency of* y_j *given* x_i, which is calculated as $\frac{a_{ij}}{t_{x_i}}$).

Certainly, the *relative frequency distribution of* X *given* Y can be dually obtained by exchanging the role of X and Y.

In the sequel, we write $f_{Y/X}$ for a relative frequency distribution of Y given X, where f_{ij} denotes its element w.r.t. the modalities x_i and y_j.

Example 2.1. Given the categorical variables $X = \{x_1, x_2\}$ representing gender (x_1 = Male and x_2 = Female) and $Y = \{y_1, y_2\}$ indicating whether a person responded positively to a survey question (y_1 = Yes and y_2 = No), we are supposed to observe the frequencies shown by Table 4.

According to Eq. (1), the total number of observations is $t_{x_1} = 30 + 20 = 50$ for "Males", $t_{x_2} = 50 + 40 = 90$ for "Females", $t_{y_1} = 30 + 50 = 80$ for "Yes", and $t_{y_2} = 20 + 40 = 60$ for "No".

Then, we can easily compute the *relative frequency distribution of* Y *given* X. For instance, the relative frequency of "Yes" given "Male" is $\frac{30}{50} = 0.6$ and the relative frequency of "No" given "Female" is $\frac{40}{90} \approx 0.444$. This yields Table 5 exhibiting the conditional distribution of Y given each modality of X.

The reverse distribution $f_{X|Y}$ can be similarly computed by exchanging the roles of rows and columns.

Table 3. Relative Frequency Distribution of Y given X.

$\mathcal{T}_{Y/X}$	y_1	$\dots$	y_j	$\dots$	y_m
x_1	$\frac{a_{11}}{t_{x_1}}$	$\dots$	$\frac{a_{1j}}{t_{x_1}}$	$\dots$	$\frac{a_{1m}}{t_{x_1}}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	$\frac{a_{i1}}{t_{x_i}}$	$\dots$	$\frac{a_{ij}}{t_{x_i}}$	$\dots$	$\frac{a_{im}}{t_{x_i}}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_n	$\frac{a_{n1}}{t_{x_n}}$	$\dots$	$\frac{a_{nj}}{t_{x_n}}$	$\dots$	$\frac{a_{nm}}{t_{x_n}}$

Table 4. Contingency Table of $X = \{\text{Male}, \text{Female}\}$ and $Y = \{\text{Yes}, \text{No}\}$.

$\mathcal{T}$	Yes	No
Male	30	20
Female	50	40

Table 5. Relative Frequency Distribution of $Y = \{\text{Yes}, \text{No}\}$ given $X = \{\text{Male}, \text{Female}\}$ (Example 2.1).

$\mathcal{T}_{Y/X}$	Yes	No
Male	0.6	0.4
Female	0.556	0.444

2.2 Generalized Fuzzy Orthopartitions

This subsection reports the concepts of intuitionistic fuzzy sets and generalized fuzzy orthopartitions.

Definition 2.2. [31] An *intuitionistic fuzzy set (IFS)* A of a universe U is a pair of functions $\mu_A : U \longrightarrow [0, 1]$ and $\nu_A : U \longrightarrow [0, 1]$ such that

$$\mu_A(u) + \nu_A(u) \leq 1,$$

for each $u \in U$.

Given $u \in U$, $\mu_A(u)$ and $\nu_A(u)$ respectively represent the *membership* and *non-membership degrees* of u to A. Moreover, the value

$$h_A(u) = 1 - (\mu_A(u) + \nu_A(u)) \tag{2}$$

is the degree of *indeterminacy* (or *uncertainty*) of u to A. Consequently, the object u belongs to A with a certain truth degree of the interval $[\mu_A(u), \mu_A(u) + h_A(u)]$.

Table 6. Intutionistic Fuzzy Sets (μ_1, ν_1) and (μ_2, ν_2) of $U = \{u_1, u_2, u_3\}$ (Example 2.4).

u	$(\mu_1(u), \nu_1(u))$	$(\mu_2(u), \nu_2(u))$
u_1	$(0.6, 0.3)$	$(0.3, 0.6)$
u_2	$(0.4, 0.4)$	$(0.5, 0.3)$
u_3	$(0.3, 0.5)$	$(0.6, 0.2)$

Let (μ_A, ν_A) be an IFS of U, the following equation is satisfied for each $u \in U$:

$$\mu_A(u) + h_A(u) = 1 - \nu_A(u). \tag{3}$$

Definition 2.3. [17] Let $O = \{(\mu_1, \nu_1), \ldots, (\mu_n, \nu_n)\}$ be a family of IFSs of U. Then, O is a generalized fuzzy orthopartition of U if and only if the following properties hold for each $u \in U$:

(i) $\sum_{i=1}^{n} \mu_i(u) \leq 1$,

(ii) $\sum_{i=1}^{n} (\mu_i(u) + h_i(u)) \geq 1$.

Observe that according to (3), Axiom (ii) can be rewritten as

$$\sum_{i=1}^{n} (1 - \nu_i(u)) \geq 1.$$

A generalized fuzzy orthopartition $O = \{(\mu_1, \nu_1), \ldots, (\mu_n, \nu_n)\}$ given by Definition 2.3, represents a partition including both vagueness and uncertainty. Indeed, as regards the meaning of Axioms (i) and (ii) of Definition 2.3, they respectively capture that the equivalence classes $C_1, \ldots, C_n$ described by $(\mu_1, \nu_1), \ldots, (\mu_n, \nu_n)$ must be disjoint and cover the initial universe[1]. Thus, for each $i \in \{1, \ldots, n\}$, its IFS (μ_i, ν_i) models an equivalence class C_i to which the objects of the initial universe belong with an unknown truth degree of a determined interval. More precisely, an element $u \in U$ belongs to C_i with a truth degree of the interval $[\mu_i(u), \mu_i(u) + h_i(u)]$.

Example 2.4. Consider the intuitionistic fuzzy sets (μ_1, ν_1) and (μ_2, ν_2) of $U = \{u_1, u_2, u_3\}$ defined by Table 6. Thus, we can immediately verify that the axioms of Definition 2.3 hold for $O = \{(\mu_1, \nu_1), (\mu_2, \nu_2)\}$.

(Axiom (i)) For each $u_i \in U$, $\mu_1(u_i) + \mu_2(u_i) \leq 1$ because $\mu_1(u_1) + \mu_2(u_1) = 0.6 + 0.3 = 0.9$, $\mu_1(u_2) + \mu_2(u_2) = 0.4 + 0.5 = 0.9$, and $\mu_1(u_3) + \mu_2(u_3) = 0.3 + 0.6 = 0.9$

(Axiom (ii)) For each $u_i \in U$, $(1 - \nu_1(u_i)) + (1 - \nu_2(u_i)) \geq 1$ because $(1 - \nu_1(u_1)) + (1 - \nu_2(u_1)) = 0.7 + 0.4 = 1.1$, $(1 - \nu_1(u_2)) + (1 - \nu_2(u_2)) = 0.6 + 0.7 = 1.3$, and $(1 - \nu_1(u_3)) + (1 - \nu_2(u_3)) = 0.5 + 0.8 = 1.3$.

Hence, O is a generalized fuzzy orthopartition of U.

On the other hand, in this article, we attach a different semantics to the IFSs of a generalized fuzzy orthopartition: they indicate sets of relative frequencies.

[1] In [22], we originally defined fuzzy orthopartitions as collections of intuitionistic fuzzy sets verifying Axioms (i) and (ii) of Definition 2.3 together with another pair of axioms. This is because the initial goal was to extend (to the fuzzy case) the notion of standard orthopartitions, which are characterized by four properties.

3 From Conditional Relative Frequency Distributions with Undefined Observations to Generalized Fuzzy Orthopartitions

This section principally reveals that generalized fuzzy orthopartitions can model evolutions of conditional relative frequency distributions, in case unknown observations are added to an initial dataset. More precisely, the following goals are achieved:

- the novel concept of conditional relative frequency distributions with undefined observations is introduced (Table 7);
- a special generalized fuzzy orthopartition is assigned to a given conditional relative frequency distribution with undefined observations (Definition 3.3 and Theorem 3.10);
- it is shown that generalized fuzzy orthopartitions and conditional relative frequencies distributions with undefined observations are equivalent models, which both capture the idea of relative frequencies characterized by uncertainty (Theorem 3.13);
- some properties of generalized fuzzy orthopartitions equivalent to conditional relative frequency distributions with undefined observations are discussed (Remarks 5 and 6 - Propositions 3.15, 3.17, and 3.19);
- finally, a generalized fuzzy orthopartition (that is equivalent to a conditional relative frequency distribution with undefined observations) is generated from a real dataset.

Let $\mathcal{T}_{Y/X}$ be a relative frequency distribution of Y given X (see Table 3 of Subsection 2.1). Then, for each $i \in \{1, \ldots, n\}$, we consider a number of α_i new and unidentified observations for $x_i \in X$ (of course, α_i is a non-negative integer). This means we cannot recognize how such observations are distributed among $y_1, \ldots, y_m$. Hence, $t_{x_i} + \alpha_i$ coincides with the sum of all frequencies related to x_i.

For each $i \in \{1, \ldots, n\}$ and for each $j \in \{1, \ldots, m\}$, we denote with k_{ij} the unknown frequency of the combination of x_i and y_j w.r.t. the new α_i observations. In symbols, we get

$$\alpha_i = k_{i1} + \ldots + k_{im}. \tag{4}$$

We can immediately observe that each k_{ij} must be an integer of the interval $[0, \alpha_i]$, namely

$$k_{ij} \in [0, \alpha_i] \cap \mathbb{Z}_0^+. \tag{5}$$

Considering all (old and new) observations for the modalities of X and Y, the relative frequency f'_{ij} of y_j given x_i is defined by the following equation:

$$f'_{ij} = \frac{a_{ij} + k_{ij}}{t_{x_i} + \alpha_i}. \tag{6}$$

Then, we use the symbol $\mathcal{T}'_{Y/X}$ to indicate the relative frequency distribution of Y given X with the $\alpha_1, \ldots, \alpha_n$ additional and unknown observations. $\mathcal{T}'_{Y/X}$ is represented by Table 7 and we say that it is a *relative frequency distribution of Y given X with $\alpha_1, \ldots, \alpha_n$ undefined observations deriving from $\mathcal{T}$*.

Shortly, we can call $\mathcal{T}'_{Y/X}$ *conditional relative frequency distribution with undefined observations*. Thus, $\mathcal{T}'_{Y/X}$ is defined by a family of parameters $\{k_{ij} \mid (i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}\}$ subject to the constraints (4) and (5).

Example 3.1. Consider Table 8, which is a simple instance of a contingency table of the variables $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$. Then, the relative frequency distribution of Y given X is represented by Table 9. Now, suppose that 5 additional observations are taken into account for x_1 and 8 for x_2, namely $\alpha_1 = 5$ and $\alpha_2 = 8$. Hence, the $\mathcal{T}'_{Y/X}$ coincides with Table 10.

The next proposition shows that the rows of $\mathcal{T}'_{Y/X}$ become relative frequency distributions for any choice of the parameters in $\{k_{ij} \mid (i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}\}$ satisfying (4) and (5).

Table 7. Relative Frequency Distribution of Y given X with $\alpha_1, \ldots, \alpha_n$ Undefined Observations deriving from $\mathcal{T}$.

$\mathcal{T}'_{Y/X}$	y_1	$\ldots$	y_j	$\ldots$	y_m
x_1	$\frac{a_{11}+k_{11}}{t_{x_1}+\alpha_1}$	$\ldots$	$\frac{a_{1j}+k_{ij}}{t_{x_1}+\alpha_1}$	$\ldots$	$\frac{a_{1m}+k_{1m}}{t_{x_1}+\alpha_1}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	$\frac{a_{i1}+k_{i1}}{t_{x_i}+\alpha_i}$	$\ldots$	$\frac{a_{ij}+k_{ij}}{t_{x_i}+\alpha_i}$	$\ldots$	$\frac{a_{im}+k_{im}}{t_{x_i}+\alpha_i}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_n	$\frac{a_{n1}+k_{n1}}{t_{x_n}+\alpha_n}$	$\ldots$	$\frac{a_{nj}+k_{nj}}{t_{x_n}+\alpha_n}$	$\ldots$	$\frac{a_{nm}+k_{nm}}{t_{x_n}+\alpha_n}$

Table 8. Contingency Table of $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$ (Example 3.1).

$\mathcal{T}$	y_1	y_2
x_1	1	2
x_2	3	4

Table 9. Relative Frequency Distribution of $Y = \{y_1, y_2\}$ given $X = \{x_1, x_2\}$ (Example 3.1).

$\mathcal{T}_{Y/X}$	y_1	y_2
x_1	$\frac{1}{3}$	$\frac{2}{3}$
x_2	$\frac{3}{7}$	$\frac{4}{7}$

Proposition 3.2. *Let $\mathcal{T}'_{Y/X}$ be a conditional relative frequency distribution with undefined observations, then $f'_{i1} + \ldots + f'_{im} = 1$, for each $i \in \{1, \ldots, n\}$.*

Proof. By (6), $f'_{i1} + \ldots + f'_{im} = \frac{a_{i1}+k_{i1}}{t_{x_i}+\alpha_i} + \ldots + \frac{a_{im}+k_{im}}{t_{x_i}+\alpha_i}$, where

$$\frac{a_{i1}+k_{i1}}{t_{x_i}+\alpha_i} + \ldots + \frac{a_{im}+k_{im}}{t_{x_i}+\alpha_i} = \frac{(a_{i1}+k_{i1}) + \ldots + (a_{im}+k_{im})}{t_{x_i}+\alpha_i} = \frac{(a_{i1}+\ldots+a_{im}) + (k_{i1}+\ldots+k_{im})}{t_{x_i}+\alpha_i}.$$

Table 10. Relative Frequency Distribution of Y given X with $\alpha_1 = 5$ and $\alpha_2 = 8$ Undefined Observations (Example 3.1).

$\mathcal{T}'_{Y/X}$	y_1	y_2
x_1	$\frac{1+k_{11}}{8}$	$\frac{2+k_{12}}{8}$
x_2	$\frac{3+k_{21}}{15}$	$\frac{4+k_{22}}{15}$

By (1), $a_{i1} + \ldots + a_{im} = t_{x_i}$; furthermore, by (4), $a_{i1} + \ldots + a_{im} = \alpha_i$. Hence, we can conclude that $f'_{i1} + \ldots + f'_{im} = \frac{t_{x_i} + \alpha_i}{t_{x_i} + \alpha_i} = 1$. □

From now on, we indicate with $\Pi_{Y/X}$ the set of all conditional relative frequency distributions arising when each k_{ij} in $\mathcal{T}'_{Y/X}$ is exactly defined. The following is its formal definition[2].

$$\Pi_{Y/X} = \{f_{Y/X} \mid f_{ij} = \frac{a_{ij} + k_{ij}}{t_{x_i} + \alpha_i} \text{ with } k_{i1} + \ldots + k_{im} = \alpha_i \text{ and } k_{ij} \in \mathbb{Z}_0^+, \text{ for each } (i,j) \in \{1,\ldots,n\} \times \{1,\ldots,m\}\}. \quad (7)$$

Moreover, we say that $\Pi_{Y/X}$ is the set of the distributions that are *consistent* with $\mathcal{T}'_{Y/X}$.

Remark 1. $\Pi_{Y/X}$ *is made of one element if and only if* $\alpha_1 = \ldots = \alpha_n = 0$. *In this case,* $\mathcal{T}'_{Y/X}$ *trivially coincides with* $\mathcal{T}_{Y/X}$.

In what follows, for each modality y_j of Y, we define a pair of functions (μ_j, ν_j) from $\{x_1, \ldots, x_n\}$ to [0,1].

Definition 3.3. We consider the mappings

$$\mu_1, \ldots, \mu_m : \{x_1, \ldots, x_n\} \longrightarrow [0,1] \quad \text{and} \quad \nu_1, \ldots, \nu_m : \{x_1, \ldots, x_n\} \longrightarrow [0,1]$$

such that for each $i \in \{1, \ldots, n\}$ and for each $j \in \{1, \ldots, m\}$,

$$\mu_j(x_i) = \frac{a_{ij}}{t_{x_i} + \alpha_i} \quad \text{and} \quad \nu_j(x_i) = 1 - \frac{a_{ij} + \alpha_i}{t_{x_i} + \alpha_i}. \quad (8)$$

Example 3.4. Let us consider $\mathcal{T}'_{Y/X}$ defined by Example 3.1 (see Table 10). Then, according to Definition 3.3, μ_1, μ_2, ν_1, and ν_2 are functions from $\{x_1, x_2\}$ to [0,1] such that

$$\mu_1(x_1) = \frac{1}{8}, \mu_1(x_2) = \frac{1}{4}, \mu_2(x_1) = \frac{1}{5}, \mu_2(x_2) = \frac{4}{15}, \nu_1(x_1) = \frac{1}{4}, \nu_1(x_2) = \frac{1}{8}, \nu_2(x_1) = \frac{4}{15}, \text{ and } \nu_2(x_2) = \frac{1}{5}.$$

Let $i \in \{1, \ldots, n\}$ and let $j \in \{1, \ldots, m\}$, by (8), we can immediately comprehend that the following chain of inequalities is true: $0 \leq \mu_j(x_i) \leq 1 - \nu_j(x_i) \leq 1$.

Example 3.5. By Example 3.4, we get $\mu_1(x_1) = \frac{1}{8}$ and $1 - \nu_1(x_1) = 1 - \frac{1}{4} = \frac{3}{4}$. As a consequence, $0 \leq \mu_1(x_1) \leq 1 - \nu_1(x_1) \leq 1$.

[2]Recall that we write $f_{Y/X}$ for a relative frequency distribution of Y given X, where f_{ij} denotes the relative frequency y_j given x_i.

Remark 2. *Let us explain the meaning of the functions introduced by Definition 3.3. Let $i \in \{1,\dots,n\}$ and let $j \in \{1,\dots,m\}$, both $\mu_j(x_i)$ and $1-\nu_j(x_i)$ can be interpreted as particular relative frequencies: according to (8), they coincide with the relative frequency of y_j given x_i when $k_{ij}=0$ and $k_{ij}=\alpha_i$, respectively.*

Example 3.6. Let us deal with $\mathcal{T}'_{Y/X}$ being defined by Example 3.1 (see Table 10). In Example 3.4, we have shown that $\mu_1(x_1)=\frac{1}{8}$ and $1-\nu_1(u_1)=\frac{3}{4}$, which correspond to the relative frequency of y_1 given x_1 when $k_{11}=0$ and $k_{11}=5$, respectively.

Remark 3. *Let $i \in \{1,\dots,n\}$ and let $j \in \{1,\dots,m\}$, the formula of $\nu_j(x_i)$ can be derived from Definition 3.3 and rewritten as $\nu_j(x_i)=\frac{t_{x_i}-a_{ij}}{t_{x_i}+\alpha_i}$. Moreover, by (1), it is easy to verify that $\nu_j(x_i)=\sum_{j^*\neq j}\frac{a_{ij^*}}{t_{x_i}+\alpha_i}$.*

Example 3.7. Let us focus on $\mathcal{T}'_{Y/X}$ being defined by Example 3.1 (see Table 10). When supposing that $k_{11}=5$ and $k_{12}=0$, $\nu_1(x_1)=\frac{a_{12}}{t_{x_1}+\alpha_1}=\frac{1}{4}$.

Although the relative frequency of y_j given x_i in Table 7 is unknown, we subsequently provide its membership range that coincides with the interval $[\mu_j(x_i), 1-\nu_j(x_i)]$. Furthermore, as already explained in Remark 2, the relative frequency of y_j given x_i equals to the endpoints $\mu_j(x_i)$ and $1-\nu_j(x_i)$, when $k_{ij}=0$ and $k_{ij}=\alpha_i$, respectively.

Theorem 3.8. *Let $f_{Y/X} \in \Pi_{Y/X}$, then*

$$\mu_j(x_i) \leq \frac{a_{ij}+k_{ij}}{t_{x_i}+\alpha_i} \leq 1-\nu_j(x_i), \tag{9}$$

for each $(i,j) \in \{1,\dots,n\}\times\{1,\dots,m\}$.

Proof. By (7), it is easy to understand that $0 \leq k_{ij} \leq \alpha_i$. Thus, the proof immediately follows from Definition 3.3. □

Example 3.9. Let us focus on $\mathcal{T}'_{Y/X}$ defined by Example 3.1 (see Table 10) and let us consider $f_{Y/X} \in \Pi_{Y/X}$ obtained by putting $k_{11}=2$, $k_{12}=3$, $k_{21}=4$, and $k_{22}=4$. In Example 3.4, we have shown that $\mu_1(x_1)=\frac{1}{8}$ and $1-\nu_1(x_1)=\frac{3}{4}$, hence $\frac{1}{8} \leq \frac{1+k_{11}}{8} \leq \frac{3}{4}$. Quickly, we can check that the inequalities (9) are also true when $(i,j) \in \{(1,2),(2,1),(2,2)\}$.

The following theorem shows that the pairs of functions $(\mu_1,\nu_1),\dots,(\mu_m,\nu_m)$ given by Definition 3.3 form a generalized fuzzy orthopartition of the universe $\{x_1,\dots,x_n\}$.

Theorem 3.10. *Let $\mu_1,\dots,\mu_m$ and $\nu_1,\dots,\nu_m$ be the mappings given by Definition 3.3, then $O_{Y/X}=\{(\mu_1,\nu_1),\dots,(\mu_m,\nu_m)\}$ is a generalized fuzzy orthopartition of $\{x_1,\dots,x_n\}$.*

Proof. First of all, it is easy to verify that let $j \in \{1,\dots,m\}$, the pair of functions (μ_j,ν_j) is an intuitionistic fuzzy set of $\{x_1,\dots,x_n\}$, according to Definition 2.2. Indeed, by Definition 3.3, let $i \in \{1,\dots,n\}$, $\mu_j(x_i)+\nu_j(x_i)=\frac{a_{ij}}{t_{x_i}+\alpha_i}+\left(1-\frac{a_{ij}+\alpha_i}{t_{x_i}+\alpha_i}\right)=\frac{t_{x_i}}{t_{x_i}+\alpha_i}$, which is obviously less than or equal to 1.

Let us prove that Axiom (i) of Definition 2.3 is satisfied by $O_{Y/X}$. By Definition 3.3, let $i \in \{1,\dots,n\}$, we get $\mu_1(x_i)+\dots\mu_m(x_i)=\frac{a_{i1}}{t_{x_i}+\alpha_i}+\dots+\frac{a_{im}}{t_{x_i}+\alpha_i}=\frac{a_{i1}+\dots+a_{im}}{t_{x_i}+\alpha_i}$. Then, by (4), $a_{i1}+\dots+a_{im}=\alpha_i$. Hence, $\mu_1(x_i)+\dots\mu_m(x_i)=\frac{\alpha_i}{t_{x_i}+\alpha_i}$, which is obviously less than or equal to 1.

Finally, let us show that Axiom (ii) of Definition 2.3 is satisfied by $O_{Y/X}$. By Definition 3.3, let $i \in \{1, \ldots, n\}$, we get $(1-\nu_1(x_i))+\ldots+(1-\nu_m(x_i)) = \frac{a_{i1}+\alpha_i}{t_{x_i}+\alpha_i}+\ldots+\frac{a_{im}+\alpha_i}{t_{x_i}+\alpha_i}$. The last member is equal to $\frac{(a_{i1}+\ldots+a_{im})+m\alpha_i}{t_{x_i}+\alpha_i}$, which coincides with $\frac{t_{x_i}+m\alpha_i}{t_{x_i}+\alpha_i}$ because $a_{i1}+\ldots+a_{im} = t_{x_i}$ from (1). Since $\frac{t_{x_i}+m\alpha_i}{t_{x_i}+\alpha_i} \geq 1$, we can conclude that $(1-\nu_1(x_i))+\ldots+(1-\nu_m(x_i)) \geq 1$. □

Table 11 represents the generalized fuzzy orthopartition $O_{Y/X} = \{(\mu_1, \nu_1), \ldots, (\mu_m, \mu_m)\}$ of $\{x_1, \ldots, x_n\}$ associated with $\mathcal{T}'_{Y/X}$ by means of Definition 3.3.

Table 11. Generalized Fuzzy Orthopartition $O_{Y/X} = \{(\mu_1, \nu_1), \ldots, (\mu_m, \nu_m)\}$ of $\{x_1, \ldots, x_n\}$ associated with $\mathcal{T}'_{Y/X}$.

$O_{Y/X}$	y_1	$\ldots$	y_j	$\ldots$	y_m
x_1	$(\mu_1(x_1), \nu_1(x_1))$	$\ldots$	$(\mu_j(x_1), \nu_j(x_1))$	$\ldots$	$(\mu_m(x_1), \nu_m(x_1))$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	$(\mu_j(x_i), \nu_j(x_i))$	$\ldots$	$(\mu_j(x_i), \nu_j(x_i))$	$\ldots$	$(\mu_m(x_i), \nu_m(x_i))$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_n	$(\mu_1(x_n), \nu_1(x_n))$	$\ldots$	$(\mu_j(x_n), \nu_j(x_n))$	$\ldots$	$(\mu_m(x_n), \nu_m(x_n))$

Notice that by (8), the values assumed by $\mu_1, \ldots, \mu_m$ and $\nu_1, \ldots, \mu_m$ are rational numbers, while generalized fuzzy orthopartitions are originally defined as pair of functions from the universe to the real interval $[0, 1]$.

Example 3.11. Let us focus on $\mathcal{T}'_{Y/X}$ defined by Example 3.1 (see Table 10). Then, the generalized fuzzy orthopartition $O_{Y/X} = \{(\mu_1, \nu_1), (\mu_2, \nu_2)\}$ of $\{x_1, x_2\}$ assigned with $\mathcal{T}'_{Y/X}$ is given by Table 12.

Table 12. Generalized Fuzzy Orthopartition $O_{Y/X} = \{(\mu_1, \nu_1), (\mu_2, \nu_2)\}$ of $\{x_1, x_2\}$ associated with $\mathcal{T}'_{Y/X}$ (Example 3.11).

$O_{Y/X}$	y_1	y_2
x_1	$\left(\frac{1}{8}, \frac{1}{4}\right)$	$\left(\frac{1}{4}, \frac{1}{8}\right)$
x_2	$\left(\frac{1}{5}, \frac{4}{15}\right)$	$\left(\frac{4}{15}, \frac{1}{5}\right)$

Remark 4. *From an algorithmic standpoint, the main computational cost in constructing a generalized fuzzy orthopartition from a conditional relative frequency distribution with undefined observations stems from Definition 3. Specifically, for each cluster $i = 1, \ldots, m$, the membership functions μ_i and ν_i must be evaluated for every element in $\{x_1, \ldots, x_n\}$. Assuming constant-time evaluation based on relative frequencies, the overall complexity is $O(n \cdot m)$. When full enumeration is infeasible, this step can be efficiently approximated using integer programming, constraint satisfaction solvers, or sampling-based methods.*

Let O be any one generalized fuzzy orthopartition of $\{x_1, \ldots, x_n\}$ so that $\mu_1, \ldots, \mu_m$ and $\nu_1, \ldots, \nu_m$ take values in $\mathbb{Q}$, then O can represent a conditional relative frequency distribution with uncertainty: the frequency of y_j given x_i is an unknown value in $[\mu_j(x_i), 1 - \nu_j(x_i)]$. In this regard, we use the symbol Π_O to denote the set of all conditional relative frequency distributions with which O could potentially coincide, once the relative frequency of y_j given x_i is known with precision, for each $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$. The following is its formal definition[3].

Definition 3.12. Let $f_{Y/X}$ be a relative frequency distribution of Y given X, then

$$f_{Y/X} \in \Pi_O \text{ if and only if } \nu_j(x_i) \leq f_{ij} \leq 1 - \nu_j(x_i) \qquad \text{for each } (i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}.$$

When $f_{Y/X} \in \Pi_O$, we say that $f_{Y/X}$ is consistent with O.

The next theorem shows that $\mathcal{T}'_{Y/X}$ and $O_{Y/X}$ have the same consistent distributions (recall that the collection $\Pi_{Y/X}$ of all distributions consistent with $\mathcal{T}'_{Y/X}$ is defined by (7)). As a consequence, $O_{Y/X}$ and $\mathcal{T}'_{Y/X}$ can considered interchangeable models: both represent the same concept of conditional relative frequency distribution with uncertainty.

Theorem 3.13. *Let $\mathcal{T}'_{Y/X}$ be a conditional relative frequency distribution with undefined observations, then* $\Pi_{O_{Y/X}} = \Pi_{Y/X}$.

Proof. It is evident that the proof immediately follows from Definitions 3.3 and 3.12. Indeed, $f_{Y/X}$ belongs to Π_O if and only if $\mu_j(x_i) \leq f_{ij} \leq 1 - \nu_j(x_i)$ for each $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$. By Definition 3.3, the latter chain of inequalities can be rewritten as $\dfrac{a_{ij}}{t_{x_i} + \alpha_i} \leq f_{ij} \leq \dfrac{a_{ij} + \alpha_i}{t_{x_i} + \alpha_i}$. Thus, $f_{ij} = \dfrac{a_{ij} + a}{t_{x_i} + \alpha_i}$ with $0 \leq a \leq \alpha_i$, which means that $f_{Y/X} \in \Pi_{Y/X}$. □

From now on, due to Theorem 3.13, we will say that $O_{Y/X}$ *is* equivalent *to* $\mathcal{T}'_{Y/X}$.

Example 3.14. Let us focus on $\mathcal{T}'_{Y/X}$ and $O_{Y/X} = \{(\mu_1, \nu_1), (\mu_2, \nu_2)\}$ defined by Tables 10 and 12, respectively. They have the same consistent relative frequency distributions.

For example, let us focus on the relative frequency distribution $f_{Y/X}$ defined by Table 13, which is determined by $k_{11} = 3$, $k_{12} = 2$, $k_{21} = 2$, and $k_{22} = 6$ and Table 10. Then, $f_{Y/X} \in \Pi_{Y/X}$. Moreover, we have $\mu_1(x_1) \leq \frac{1}{2} \leq 1 - \nu_1(x_1)$,

Table 13. Relative Frequency Distribution of Y given X obtained by putting $k_{11} = 3$, $k_{12} = 2$, $k_{21} = 2$, and $k_{22} = 6$ in Table 10 (Example 3.14).

$f_{Y/X}$	y_1	y_2
x_1	$\frac{1}{2}$	$\frac{1}{2}$
x_2	$\frac{1}{3}$	$\frac{2}{3}$

[3]This concept is syntactically equivalent to that of consistent Ruspini partitions when we attach the standard meaning to generalized fuzzy orthopartitions (see [22]).

$\mu_2(x_1) \leq \frac{1}{2} \leq 1 - \nu_2(x_1)$, $\mu_1(x_2) \leq \frac{1}{3} \leq 1 - \nu_1(x_2)$, and $\mu_2(x_2) \leq \frac{2}{3} \leq 1 - \nu_2(x_2)$. Consequently, by Definition 3.12, $f_{Y/X} \in \Pi_{O_{Y/X}}$.

Vice-versa, we can take into account the relative frequency distribution $f^*_{Y/X}$ given by Table 14, which belongs to $\Pi_{O_{Y/X}}$. It is easy to check that $f^*_{Y/X}$ is consistent with $\mathcal{T}'_{Y/X}$, in fact it can be obtained by $k_{11} = 4$, $k_{12} = 1$, $k_{21} = 5$,

Table 14. Relative Frequency Distribution of Y given X satisfying Condition 12 (Example 3.14).

$f^*_{Y/X}$	y_1	y_2
x_1	$\frac{5}{8}$	$\frac{3}{8}$
x_2	$\frac{8}{15}$	$\frac{7}{15}$

and $k_{22} = 3$ in Table 10.

In the sequel, we discuss some properties of generalized fuzzy orthopartitions assigned to conditional relative frequency distributions with undefined observations. Thus, we are considering $O_{Y/X} = \{(\mu_1, \nu_1), \ldots, (\mu_n, \nu_n)\}$ associated with $\mathcal{T}'_{Y/X}$.

Proposition 3.15. *For each* $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$,

$$(1 - \nu_j(x_i)) + \sum_{j^* \neq j} \mu_{j^*}(x_i) = 1. \tag{10}$$

Proof. Let $i \in \{1, \ldots, n\}$ and let $j \in \{1, \ldots, m\}$, by Definition 3.3,

$$(1 - \nu_j(x_i)) + \sum_{j^* \neq j} \mu_{j^*}(x_i) = \frac{a_{ij} + \alpha_i}{t_{x_i} + \alpha_i} + \sum_{j^* \neq j} \frac{a_{ij^*}}{t_{x_i} + \alpha_i}. \tag{11}$$

Of course, the second member of Equation (11) is equal to $\dfrac{\sum_{j=1}^{n} a_{ij} + \alpha_i}{t_{x_i} + \alpha_i}$. Moreover, by (1), we get $\sum_{j=1}^{n} a_{ij} = t_{x_i}$. Hence, we can conclude that

$$(1 - \nu_j(x_i)) - \sum_{j^* \neq j} \mu_{j^*}(x_i) = \frac{t_{x_i} + \alpha_i}{t_{x_i} + \alpha_i} = 1.$$

□

Example 3.16. Let us consider the contingency table of $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$ given by Table 15. Then, $t_{x_1} = 8$ and $t_{x_2} = 9$. Supposing that $\alpha_1 = 6$ and $\alpha_2 = 7$, $\mathcal{T}'_{Y/X}$ is defined by Table 16 and the generalized fuzzy orthopartition $O_{Y/X}$ equivalent to $\mathcal{T}'_{Y/X}$ is represented by Table 17.

According to Proposition 3.15, (10) is satisfied for each pair of indices $(i, j) \in \{1, 2\} \times \{1, 2, 3\}$: $(1 - \nu_1(x_1)) + \mu_2(x_1) + \mu_3(x_1) = \frac{1}{2} + \frac{5}{14} + \frac{1}{7} = 1$; $(1 - \nu_2(x_1)) + \mu_1(x_1) + \mu_3(x_1) = \frac{11}{14} + \frac{1}{14} + \frac{1}{7} = 1$; $(1 - \nu_3(x_1)) + \mu_1(x_1) + \mu_2(x_1) =$

Table 15. Contingency Table of $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$ (Example 3.16).

$\mathcal{T}$	y_1	y_2	y_3
x_1	1	5	2
x_2	4	1	4

Table 16. Relative Frequency Distribution of Y given X with $\alpha_1 = 6$ and $\alpha_2 = 7$ (Example 3.16).

$\mathcal{T}'_{Y/X}$	y_1	y_2	y_3
x_1	$\frac{1+k_{11}}{14}$	$\frac{5+k_{12}}{14}$	$\frac{2+k_{13}}{14}$
x_2	$\frac{4+k_{21}}{16}$	$\frac{1+k_{22}}{16}$	$\frac{4+k_{23}}{16}$

Table 17. Generalized Fuzzy Orthopartition $O_{Y/X} = \{(\mu_1, \nu_1), (\mu_2, \nu_2), (\mu_3, \nu_3)\}$ of $\{x_1, x_2\}$ equivalent to $\mathcal{T}'_{Y/X}$ (Example 3.16).

$O_{Y/X}$	y_1	y_2	y_3
x_1	$\left(\frac{1}{14}, \frac{1}{2}\right)$	$\left(\frac{5}{14}, \frac{3}{14}\right)$	$\left(\frac{1}{7}, \frac{3}{7}\right)$
x_2	$\left(\frac{1}{4}, \frac{5}{16}\right)$	$\left(\frac{1}{16}, \frac{1}{2}\right)$	$\left(\frac{1}{4}, \frac{5}{16}\right)$

$\frac{4}{7} + \frac{1}{14} + \frac{5}{14} = 1$; $(1 - \nu_1(x_2)) + \mu_2(x_2) + \mu_3(x_2) = \frac{11}{16} + \frac{1}{16} + \frac{1}{4} = 1$; $(1 - \nu_2(x_2)) + \mu_1(x_2) + \mu_3(x_2) = \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1$; $(1 - \nu_3(x_2)) + \mu_1(x_2) + \mu_2(x_2) = \frac{11}{16} + \frac{1}{4} + \frac{1}{16} = 1$.

Proposition 3.15 is necessary so that $\Pi_{O_{Y/X}}$ contains any distribution that can be obtained from $\mathcal{T}'_{Y/X}$ as follows [4]: let $i \in \{1, \ldots, n\}$, we fix $j^* \in \{1, \ldots, m\}$ and we put

$$k_{ij^*} = \alpha_i \quad \text{and} \quad k_{ij} = 0, \quad \text{for each } j \neq j^*. \tag{12}$$

Indeed, if $f_{Y/X}$ is a distribution of this type, then for each $i \in \{1, \ldots, n\}$, there exists $j^* \in \{1, \ldots, m\}$ such that

$$f_{ij^*} = \frac{a_{ij^*} + \alpha_i}{t_{x_i} + \alpha_i} \quad \text{and} \quad f_{ij} = \frac{a_{ij}}{t_{x_i} + \alpha_i}$$

[4]Recall that $\Pi_{O_{Y/X}}$ equals $\Pi_{Y/X}$ from Theorem 3.13.

from (7) and (12); hence, by Definition 3.3, such relative frequencies can be rewritten as $f_{ij^*} = 1 - \nu_{j^*}(x_i)$ and $f_{ij} = \mu_j(x_i)$. Then, $\mu_1, \ldots, \mu_m$ and $\nu_1, \ldots, \nu_m$ must be defined so that $\{1 - \nu_{j^*}(x_i)\} \cup \{\mu_j(x_i) \mid j \neq j^*\}$ can be viewed as a relative frequency distribution, and for this reason, $O_{Y/X}$ needs to satisfy (10) for each $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$.

Proposition 3.17. *For each* $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$,

$$h_j(x_i) = \frac{\alpha_i}{t_{x_i} + \alpha_i}. \tag{13}$$

Proof. Given $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, $h_j(x_1) = 1 - (\mu_j(x_i) + \nu_j(x_i))$ from (2). Moreover, $\mu_j(x_i) = \dfrac{a_{ij}}{t_{x_i} + \alpha_i}$ and $\nu_j(x_i) = 1 - \dfrac{a_{ij} + \alpha_i}{t_{x_i} + \alpha_i}$ from Definition 3.3. Then, we get

$$h_j(x_i) = 1 - \left(\frac{a_{ij}}{t_{x_i} + \alpha_i} + 1 - \frac{a_{ij} + \alpha_i}{t_{x_i} + \alpha_i} \right) = \frac{\alpha_i}{t_{x_i} + \alpha_i}.$$

□

Remark 5. *According to Proposition 3.17, the degrees of uncertainty*[5] *related to the modality* $x_i \in X$ *are equal to each other (in symbols,* $h_1(x_i) = \ldots = h_m(x_i)$*). This is because* $h_1(x_i), \ldots, h_m(x_i)$ *respectively depend on the parameters* $k_{i1}, \ldots, k_{im}$ *(recall that* $O_{Y/X}$ *is constructed by starting from* $\mathcal{T}'_{Y/X}$*), which all have the set* $[0, \alpha_i] \cap \mathbb{Z}_0^+$ *as range of possible values.*

Example 3.18. Let us go back to $O_{Y/X}$ defined by Table 17. Let $(i, j) = (1, 1)$, we get $h_1(x_1) = 1 - (\mu_1(x_1) + \nu_1(x_1)) = 1 - \left(\frac{1}{14} + \frac{1}{2} \right) = \frac{3}{7}$ and $\frac{\alpha_1}{t_{x_1} + \alpha_1} = \frac{6}{8 + 6} = \frac{3}{7}$ and $h_2(x_1) = h_3(x_1) = \frac{3}{7}$ (k_{11}, k_{12}, and k_{13} are unknown integers between 0 and 6). Furthermore, it is easy to verify that $h_1(x_2) = h_2(x_2) = h_3(x_2) = \frac{7}{16}$ (k_{21}, k_{22}, and k_{23} are unknown integer between 0 and 7).

Proposition 3.19. *Let* $O = \{(\mu_1, \nu_1) \ldots, (\mu_m, \nu_m)\}$ *be a generalized fuzzy orthopartition of* $\{x_1, \ldots, x_n\}$ *such that Eq.* (10) *holds for each* $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$*. Then,*

$$h_j(x_i) = 1 - (\mu_1(x_i) + \ldots + \mu_m(x_i)), \quad \textit{for each } (i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}.$$

Proof. Let $i \in \{1, \ldots, n\}$ and let $j \in \{1, \ldots, m\}$, then $h_j(x_i) = 1 - (\mu_j(x_i) + \nu_j(x_i))$ from (2). By hypothesis, Eq. (10) is true; hence, $\nu_j(x_i) = \sum_{j' \neq j} \mu_{j'}(x_i)$. Then, $h_j(x_i) = 1 - (\mu_j(x_i) + \nu_j(x_i)) = 1 - (\mu_j(x_i) + \sum_{j' \neq j} \mu_{j'}(x_i)) = 1 - (\mu_1(x_i) + \ldots + \mu_m(x_i))$. □

In order to streamline the notation, we set $h_i = h_j(x_i)$ for each $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$, since $h_1(x_i) = \ldots = h_m(x_i)$ due to Remark 5.

Remark 6. *Given* $i \in \{1, \ldots, n\}$*, the number of possible values that the parameters* $k_{i1}, \ldots, k_{im}$ *in* $\mathcal{T}'_{Y/X}$ *can assume depends on the integer* α_i*. Indeed, the larger* α_i *is, the more numerous the possible choices are for* $k_{i1}, \ldots, k_{im}$*. As a consequence, the greater* $\alpha_1, \ldots, \alpha_n$ *are, the bigger the set* $\Pi_{Y/X}$ *is. Also, due to Proposition 3.17,* h_i *depends on* α_i[6]*: as the increase of* α_i*, the greater* $h_j(x_i)$ *is. Then, we can deduce that the greater the values* $h_1, \ldots, h_n$ *are, the bigger the set* $\Pi_{Y/X}$ *is. For this reason,* $\dfrac{h_1 + \ldots + h_n}{n} \in [0, 1)$ *can be considered as a* measure of entropy *capturing the quantity*

[5]When (μ_j, ν_j) is an IFS of $\{x_1, \ldots, x_n\}$, $h_j(x_i)$ is usually called degree of indeterminacy or uncertainty, as reported in Subsection 2.2.
[6]The function $f(x) = \dfrac{c}{x + c}$, where c is a constant and $x \geq 0$, is increasing.

$(\mathcal{T}_1)'_{Y/X}$	y_1	y_2
x_1	$\frac{1+k_{11}}{3}$	$\frac{2+k_{12}}{3}$
x_2	$\frac{3+k_{21}}{8}$	$\frac{4+k_{22}}{8}$

$(\mathcal{T}_2)'_{Y/X}$	y_1	y_2
x_1	$\frac{1+k_{11}}{4}$	$\frac{2+k_{12}}{4}$
x_2	$\frac{3+k_{21}}{8}$	$\frac{4+k_{22}}{8}$

Fig. 2. Conditional Relative Frequency Distributions with Undefined Observations $(\mathcal{T}_1)'_{Y/X}$ and $(\mathcal{T}_2)'_{Y/X}$ (Example 3.20).

$(O_1)_{Y/X}$	y_1	y_2
x_1	$\left(\frac{1}{3}, \frac{2}{3}\right)$	$\left(\frac{2}{3}, \frac{1}{3}\right)$
x_2	$\left(\frac{3}{8}, \frac{1}{2}\right)$	$\left(\frac{1}{2}, \frac{3}{8}\right)$

$(O_2)_{Y/X}$	y_1	y_2
x_1	$\left(\frac{1}{4}, \frac{1}{2}\right)$	$\left(\frac{1}{2}, \frac{1}{4}\right)$
x_2	$\left(\frac{3}{8}, \frac{1}{2}\right)$	$\left(\frac{1}{2}, \frac{3}{8}\right)$

Fig. 3. Generalized Fuzzy Orthopartitions $(O_1)_{Y/X}$ and $(O_2)_{Y/X}$ (Example 3.20).

of uncertainty contained in $\mathcal{T}'_{Y/X}$*. Thus, as the increase of* $\frac{h_1 + \ldots + h_n}{n}$*, the bigger* $\Pi_{Y/X}$ *becomes, hence the number of distributions with which* $\mathcal{T}'_{Y/X}$ *could coincide (once the uncertainty is solved from* $\mathcal{T}'_{Y/X}$ [7]*) increases. Furthermore, by Remark 1,* $\frac{h_1 + \ldots + h_n}{n}$ *is equal to 0 if and only if* $\mathcal{T}'_{Y/X}$ *is identified with a specific conditional relative frequency distribution.*

Example 3.20. Consider the conditional relative frequency distributions with undefined observations $(\mathcal{T}_1)'_{Y/X}$ and $(\mathcal{T}_2)'_{Y/X}$, which are represented by the tables in Figure 2. Both are obtained by Table 8: $(\mathcal{T}_1)'_{Y/X}$ when $\alpha_1 = 0$ and $\alpha_2 = 1$ and $(\mathcal{T}_2)'_{Y/X}$ when $\alpha_1 = \alpha_2 = 1$. Then, their generalized fuzzy orthopartitions $(O_1)_{Y/X}$ and $(O_2)_{Y/X}$ are represented in Figure 3. It is easy to check that $\frac{h_1 + h_2}{2} = \frac{0 + 1/8}{2} = \frac{1}{16}$ for $(O_1)_{Y/X}$ and $\frac{h_1 + h_2}{2} = \frac{1/4 + 1/8}{2} = \frac{3}{16}$ for $(O_2)_{Y/X}$. So, in line with Remark 6, $(\mathcal{T}_1)'_{Y/X}$ has less consistent distributions than $(\mathcal{T}_2)'_{Y/X}$: notice that the distributions consistent with $(\mathcal{T}_1)'_{Y/X}$ are given by $(k_{11}, k_{12}, k_{21}, k_{22}) \in \{(0, 0, 1, 0), (0, 0, 0, 1)\}$; while, those consistent with $(\mathcal{T}_2)'_{Y/X}$ are given by $(k_{11}, k_{12}, k_{21}, k_{22}) \in \{(1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0), (0, 1, 0, 1)\}$.

Up to this point, we have assigned a generalized fuzzy orthopartition $O_{Y/X}$ to a relative frequency distribution $\mathcal{T}_{Y/X}$, in the presence of a given number of undefined observations. In turn, $\mathcal{T}_{Y/X}$ is derived from a contingency table $\mathcal{T}$ as explained in Subsection 2.1. Of course, there exist infinitely many contingency tables generating $\mathcal{T}_{Y/X}$; consequently, we can immediately deduce that $O_{Y/X}$ corresponds to exactly one conditional relative frequency

[7] The uncertainty is solved from $\mathcal{T}'_{Y/X}$ when we can assign a precise value to each k_{ij} in $\mathcal{T}'_{Y/X}$.

distribution and an infinite class of contingency tables. The main results of this section are schematized in Figure 4 [8].

Moreover, we can generate dual results by changing the role of the variables X and Y, and so, we can obtain generalized fuzzy orthopartitions of $\{y_1, \ldots, y_m\}$ from relative frequency distributions of X given Y.

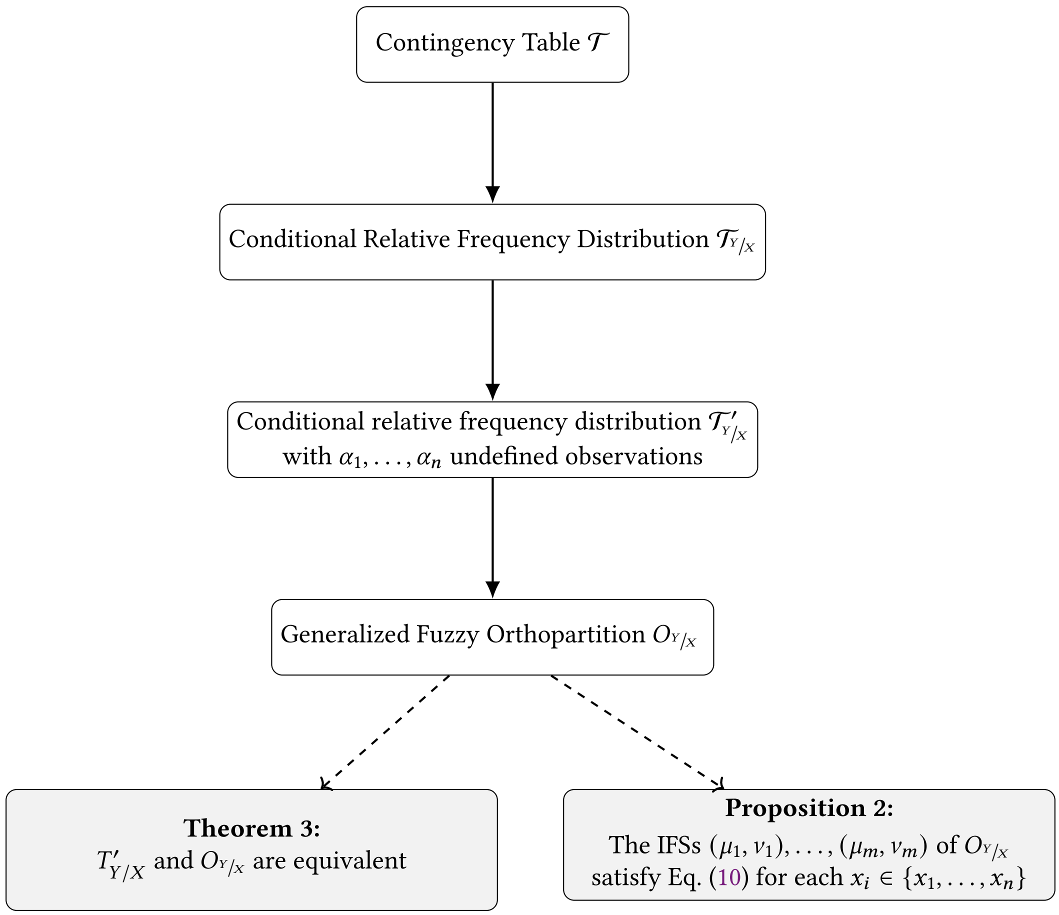

Fig. 4. Logical structure of Section 3 results: from Contingency Tables to Generalized Fuzzy Orthopartitions.

Illustrative example. We present an example of generalized fuzzy orthopartition deriving from a conditional relative frequency distribution, which is generated by a real dataset. We deal with the contingency table in Figure 5 that expresses the relationship between the categorical variables X = *"Italian Region"* and Y = *"Tourist Attraction"*. Then, Y has the modalities y_1, y_2, and y_3, where

y_1 = *"museum, non-profit gallery and/or collection"*,
y_2 = *"archaeological area or park"*,
y_3 = *"monument or monumental complex"*.

[8] Recall that a conditional relative frequency distribution with undefined observations and a generalized fuzzy orthopartition are equivalent when they represent the same class of conditional relative frequency distributions.

The modalities of X are indicated with the symbols $x_1, \ldots, x_{21}$, where x_1 = *"Piemonte"*, x_2 = *"Valle d' Aosta"*, x_4 = *"Lombardia"*, x_5 = *"Provincia Autonoma di Bolzano"*, x_6 = *"Provincia Autonoma di Trento"*, x_7 = *"Veneto"*, x_8 = *"Friuli-Venezia Giulia"*, x_9 = *"Emilia Romagna"*, x_{10} = *"Toscana"*, x_{11} = *"Umbria"*, x_{12} = *"Marche"*, x_{13} = *"Lazio"*, x_{14} = *"Abruzzo"*, x_{15} = *"Molise"*, x_{16} = *"Campania"*, x_{17} = *"Puglia"*, x_{18} = *"Basilicata"*, x_{19} = *"Calabria"*, x_{20} = *"Sicilia"*, and x_{21} = *"Sardegna"*.

Given an Italian region x_i and a tourist attraction y_j, the cell related to the i-th row and the j-th column in the table $\mathcal{T}$ of Figure 5 contains the total number of visitors of the tourist attraction y_j in the region x_i, in the year 2015. For example, taking into account the modalities x_1 (*Piemonte*) and y_2 (*archaeological area or park*), $\mathcal{T}$ shows that 26306 people visited an archaeological area or park in Piemonte, in the year 2015. $\mathcal{T}$ carries real data, which have been detected by the *Italian National Institute of Statistics (ISTAT)* [9]. A second table $\mathcal{T}_{Y/X}$ in Figure 5 represents the relative frequency distribution of Y given X. For instance, the cell value assigned to the modalities x_1 (*Piemonte*) and y_1 (*museums, non-profit galleries and/or collections*) is $\frac{4998069}{6692925} = 0.75$, where

- 4998069 is the number of visitors at *museums, non-profit galleries and/or collections* in *Piemonte*, in the year 2015;
- 6692925 is the total number of tourist attractions in *Piemonte*, during the same year.

We suppose 50000 additional and unknown observations are made for each Italian region. Then, according to this article notation, $\alpha_1 = \ldots = \alpha_{21} = 50000$. This assumption could be required, for example, in case 50000 new tickets are sold in every region and each of them allows a visit to only one attraction that cannot be predicted.

Thus, generalized relative frequencies are determined by (6). As an example, the relative frequency of y_1 given x_1 is

$$f'_{11} = \frac{4998069 + k_{11}}{6692925 + 50000} = \frac{4998069 + k_{11}}{6742925},$$

where k_{11} is an unknown integer of the interval [0, 50000]. Figure 6 defines the generalized fuzzy orthopartition $O_{Y/X} = \{(\mu_1, \nu_1), (\mu_2, \nu_2), (\mu_3, \nu_3)\}$ of $\{x_1, \ldots, x_{21}\}$, which is equivalent to $\mathcal{T}'_{Y/X}$ and models the relationship between X = *Italian Region* and Y = *Tourist Attraction*, in presence of the 50000 uncertain observations for each Italian region.

In line with Theorem 3.8, $O_{Y/X}$ provides the membership intervals of relative frequencies consistent with $\mathcal{T}'_{Y/X}$. For istance, focusing on the modalities x_1 and y_1, we have $(\mu_1(x_1), \nu_1(x_1)) = (0.73, 0.25)$, where $\mu_1(x_1)$ and $\nu_1(x_1)$ are calculated by means of Definition 3.3:

- $\mu_1(x_1) = \frac{a_{11}}{t_{x_1} + \alpha_1} = \frac{4998069}{6692925 + 50000} = \frac{4998069}{6742925} = 0.73;$
- $\nu_1(x_1) = 1 - \frac{a_{11} + \alpha_1}{t_{x_1} + \alpha_1} = \frac{4998069 + 50000}{6692925 + 50000} = 1 - \frac{5058069}{6742925} = 1 - 0.75 = 0.25.$

Therefore, the relative frequency of y_1 given x_1 (i.e. the relative frequency of the visitors of a *museum, non-profit gallery and/or collection* in Piemonte) is a value between 0.73 and $1 - 0.25 = 0.75$.

4 From Generalized Fuzzy Orthopartition to Conditional Relative Frequency Distributions with Undefined Observations

This section addresses the issue of determining the class of conditional relative frequency distributions with undefined observations that are equivalent to a given generalized fuzzy orthopartition $O = \{(\mu_1, \nu_1), \ldots, (\mu_m, \nu_m)\}$ of $\{x_1, \ldots, x_n\}$ satisfying the condition exhibited by Proposition 3.15. In this case, as previously mentioned, the IFSs of generalized fuzzy orthopartitions assume their values in $\mathbb{Q}$. Moreover, it is important to underline that O is equivalent to a conditional relative frequency distribution with undefined observations $\mathcal{T}'_{Y/X}$, when its elements can

[9]Such data are available on the ISTAT website (https://www.istat.it/).

$\mathcal{T}$	y_1	y_2	y_3
x_1	4998069	26306	1668550
x_2	125307	201740	533644
x_3	1453374	15507	166695
x_4	6784121	405364	1676724
x_5	1343473	1650	177099
x_6	1688139	31531	189914
x_7	6789439	20965	2007455
x_8	1135959	85864	1289949
x_9	4244013	58591	1313160
x_{10}	13921448	229553	8641349
x_{11}	1042130	65928	558033
x_{12}	1678911	27491	150194
x_{13}	5079100	2404077	17120821
x_{14}	371555	58417	88270
x_{15}	144349	35201	17870
x_{16}	3886028	4205730	2101808
x_{17}	682142	41028	507687
x_{18}	304278	89945	2250
x_{19}	1499712	87968	176341
x_{20}	1526186	2942248	769923
x_{21}	900270	517444	258978

$\mathcal{T}_{Y/X}$	y_1	y_2	y_3
x_1	0.75	0	0.25
x_2	0.15	0.23	0.62
x_3	0.89	0.01	0.1
x_4	0.76	0.05	0.19
x_5	0.9	0	0.1
x_6	0.9	0	0.1
x_7	0.77	0	0.23
x_8	0.46	0.03	0.51
x_9	0.75	0.01	0.24
x_{10}	0.61	0.01	0.38
x_{11}	0.62	0.04	0.34
x_{12}	0.9	0.02	0.08
x_{13}	0.2	0.1	0.7
x_{14}	0.72	0.11	0.17
x_{15}	0.73	0.17	0.1
x_{16}	0.39	0.41	0.2
x_{17}	0.56	0.03	0.41
x_{18}	0.77	0.23	0
x_{19}	0.85	0.05	0.1
x_{20}	0.3	0.56	0.14
x_{21}	0.54	0.31	0.15

Fig. 5. Contingency Table of X and Y and Relative Frequency Distribution of Y given X, where $Y =$*"Tourist Attraction"* and $X =$*"Italian Region"*.

be determined from $\mathcal{T}'_{Y/X}$ by means of Definition 3.3. The main results achieved are summarized in the following points.

- A procedure is described to construct a class $\mathcal{F}_O$ of contingency tables having the same conditional relative frequency distribution (up to and including Proposition 4.2).

$O_{Y/X}$	y_1	y_2	y_3
x_1	(0.73, 0.25)	(0.004, 0.985)	(0.24, 0.75)
x_2	(0.13,0.79)	(0.21,0.7)	(0.56,0.45)
x_3	(0.84,0.11)	(0.01,0.85)	(0.097,0.857)
x_4	(0.75, 0.24)	(0.04, 0.95)	(0.18, 0.81)
x_5	(0.83, 0.11)	(0.001, 0.95)	(0.11, 0.84)
x_6	(0.84, 0.12)	(0.015, 0.944)	(0.095, 0.865)
x_7	(0.76, 0.23)	(0.002, 0.989)	(0.22, 0.77)
x_8	(0.43, 0,54)	(0.03, 0.94)	(0.5, 0.5)
x_9	(0.74, 0.25)	(0.01, 0.98)	(0.23, 0.76)
x_{10}	(0.6, 0.39)	(0.01, 0.987)	(0.37, 0.62)
x_{11}	(0.59, 0.36)	(0.03, 0.92)	(0.31, 0.64)
x_{12}	(0.86, 0.1)	(0.014, 0.95)	(0.07, 0.89)
x_{13}	(0.2, 0.8)	(0.09, 0.9)	(0.69, 0.31)
x_{14}	(0.62, 0.25)	(0.09, 0.77)	(0.14, 0,72)
x_{15}	(0.52, 0.2)	(0.12, 0.59)	(0.06, 0.65)
x_{16}	(0.37, 0.62)	(0.4, 0.59)	(0.20, 0.79)
x_{17}	(0.52, 0.42)	(0.03, 0.91)	(0.38, 0.56)
x_{18}	(0.63, 0.2)	(0.18, 0,65)	(0.004, 0.83)
x_{19}	(0.81, 0.15)	(0.04, 0.91)	(0.09, 0.87)
x_{20}	(0.28, 0.7)	(0.55, 0.44)	(0.14, 0.85)
x_{21}	(0.51, 0.45)	(0.29, 0.66)	(0.14, 0.81)

Fig. 6. Generalized Fuzzy Orthopartition $O_{Y/X} = \{(\mu_1, \nu_1), (\mu_2, \nu_2), (\mu_3, \nu_3)\}$ associated with $\mathcal{T}'_{Y/X}$.

- For each contingency table in $\mathcal{F}_O$, the integers $\alpha_1, \dots, \alpha_n$ are determined so that the corresponding conditional relative frequency distribution with $\alpha_1, \dots, \alpha_n$ observations is equivalent to O (Theorem 4.5).
- It is shown that all the distributions equivalent to O can be derived from the elements of $\mathcal{F}_O$ (Theorem 4.7).
- A characterization for generalized fuzzy orthopartitions coinciding with conditional relative frequency distributions with undefined observations is provided (Theorem 4.9).

- The results of Sections 3 and 4 are connected: a conditional relative frequency distribution with undefined observations $\mathcal{T}'_{Y/X}$ is obtained from a specific contingency table of $\mathcal{F}_{O_{Y/X}}$ (Theorem 4.10).

In the sequel, given a generalized fuzzy orthopartition $O = \{(\mu_1, \nu_1), \ldots, (\mu_m, \nu_m)\}$ of $\{x_1, \ldots, x_n\}$ such that Eq. (10) holds for each $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$, we construct a family $\mathcal{F}_O$ of contingency tables of X and Y by following four steps.

(1) For each $i \in \{1, \ldots, n\}$, we write $\mu_1(x_i), \ldots, \mu_m(x_i)$ and h_i as irreducible fractions (this is possible because we have supposed that $\mu_1(x_i), \ldots, \mu_m(x_i)$ and h_i belong to $\mathbb{Q}$):

$$\mu_1(x_i) = \frac{q_{i1}}{p_{i1}}, \ldots, \mu_m(x_i) = \frac{q_{im}}{p_{im}} \quad \text{and} \quad h_i = \frac{s_i}{r_i}. \tag{14}$$

Certainly, $q_{i1}, \ldots, q_{im}, s_i$ are non negative integers and $p_{i1}, \ldots, p_{im}, r_i$ are positive integers such that $q_{ij} \leq p_{ij}$ for each $j \in \{1, \ldots, m\}$ and $s_i \leq r_i$.

(2) We find out the equivalent fractions to $\frac{q_{i1}}{p_{i1}}, \ldots, \frac{q_{im}}{p_{im}}$ and $\frac{s_i}{r_i}$, which have the lowest common multiple of $p_{i1}, \ldots, p_{im}$, and r_i as denominator. Thus, we get

$$\mu_1(x_i) = \frac{q'_{i1}}{t_i}, \ldots, \mu_m(x_i) = \frac{q'_{im}}{t_i} \quad \text{and} \quad h_i = \frac{s'_i}{t_i}, \tag{15}$$

where t_i is the lowest common multiple of $p_{i1}, \ldots, p_{im}$, and r_i.

(3) We consider Table 18 indicated with $\mathcal{T}_O$, where the cell value related to the i-row and the j-column is $\mu_j(x_i) \cdot t_i$. Notice that by (15), $\mu_j(x_i) \cdot t_i$ equals q'_{ij}, which is obviously a non-negative integer.

Table 18. Contingency Table of X and Y assigned to O.

$\mathcal{T}_O$	y_1	$\ldots$	y_j	$\ldots$	y_m
x_1	$\mu_1(x_1) \cdot t_1$	$\ldots$	$\mu_j(x_1) \cdot t_1$	$\ldots$	$\mu_m(x_1) \cdot t_1$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	$\mu_1(x_i) \cdot t_i$	$\ldots$	$\mu_j(x_i) \cdot t_i$	$\ldots$	$\mu_m(x_i) \cdot t_i$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_n	$\mu_1(x_n) \cdot t_n$	$\ldots$	$\mu_j(x_n) \cdot t_n$	$\ldots$	$\mu_m(x_n) \cdot t_n$

(4) We define $\mathcal{F}_O$ as the collection of all tables arising from multiplying each element of the i-th row of $\mathcal{T}_O$ by a positive integer a_i. Therefore, let $a_1, \ldots, a_n$ be positive integers, Table 19 denoted with $\mathcal{T}_O(a_1, \ldots, a_n)$ belongs to $\mathcal{F}_O$. In symbols, $\mathcal{F}_O$ is defined as follows:

$$\mathcal{F}_O = \{\mathcal{T}_O(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \in \mathbb{Z}^+\}, \tag{16}$$

where $\mathbb{Z}^+$ is the set of positive integers.

Example 4.1. Let $O = \{(\mu_1, \nu_1), (\mu_2, \nu_2), (\mu_3, \nu_3)\}$ be the generalized fuzzy orthopartition of $\{x_1, x_2\}$ defined by Table 17. Let us retrace the steps of the procedure explained above to construct $\mathcal{F}_O$.

(1) Firstly, we can notice that $\mu_1(x_1), \mu_2(x_1), \mu_3(x_1)$ and $\mu_1(x_2), \mu_2(x_2), \mu_3(x_2)$ are already written as irreducible fractions. Also, as exhibited by Example 3.18, $h_1 = \frac{3}{7}$ and $h_2 = \frac{7}{16}$.

Table 19. Contingency Table of X and Y assigned to O, where $a_1, \ldots, a_n$ are positive integers.

$\mathcal{T}_O(a_1, \ldots, a_n)$	y_1	$\ldots$	y_j	$\ldots$	y_m
x_1	$\mu_1(x_1) \cdot t_1 \cdot a_1$	$\ldots$	$\mu_j(x_1) \cdot t_1 \cdot a_1$	$\ldots$	$\mu_m(x_1) \cdot t_1 \cdot a_1$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	$\mu_1(x_i) \cdot t_i \cdot a_i$	$\ldots$	$\mu_j(x_i) \cdot t_i \cdot a_i$	$\ldots$	$\mu_m(x_i) \cdot t_i \cdot a_i$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_n	$\mu_1(x_n) \cdot t_n \cdot a_n$	$\ldots$	$\mu_j(x_n) \cdot t_n \cdot a_n$	$\ldots$	$\mu_m(x_n) \cdot t_n \cdot a_n$

(2) In line with the notation adopted before (see Step 2), $p_{11} = 14$, $p_{12} = 14$, $p_{13} = 7$, and $r_1 = 7$. Moreover, $p_{21} = 4$, $p_{22} = 16$, $p_{23} = 4$, and $r_2 = 16$.
Then, $t_1 = 14$ and $t_2 = 16$, which are respectively the lowest common multiple of $\{p_{11}, p_{12}, p_{13}, r_1\}$ and $\{p_{21}, p_{22}, p_{23}, r_2\}$. Consequently, we get

$$\mu_1(x_1) = \frac{1}{14}, \mu_2(x_1) = \frac{5}{14}, \mu_3(x_1) = \frac{2}{14}, h_1(x_1) = h_2(x_1) = h_3(x_1) = \frac{6}{14}$$

and

$$\mu_1(x_2) = \frac{4}{16}, \mu_2(x_2) = \frac{1}{16}, \mu_3(x_2) = \frac{4}{16}, h_1(x_2) = h_2(x_2) = h_3(x_2) = \frac{7}{16}.$$

(3) Therefore, $\mathcal{T}_O$ coincides with Table 15. For example, we can view that the element in its first row and its second column is $\mu_2(x_1) \cdot t_1 = \frac{1}{14} \cdot 14 = 1$.

(4) By (16), $\mathcal{F}_O = \{\mathcal{T}_O(a_1, a_2) \mid a_1, a_2 \in \mathbb{Z}^+\}$. Concretely, each other object of $\mathcal{F}_O$ can be determined by multiplying the elements of the first and the second rows of Table 15 by the positive integers a_1 and a_2, respectively. For example, let $a_1 = a_2 = 2$, $\mathcal{T}_O(a_1, a_2)$ coincides with Table 20.

Table 20. $\mathcal{T}_O(a_1, a_2)$ of $\mathcal{F}_O$ with $a_1 = a_2 = 2$ (Example 4.1).

$\mathcal{T}_O(2, 2)$	y_1	y_2	y_3
x_1	2	10	4
x_2	8	2	8

We list some simple observations characterizing the class $\mathcal{F}_O$.

Remark 7. *Let* $\mathcal{T}_O(a_1, \ldots, a_n) \in \mathcal{F}_O$,

$$\mathcal{T}_O(a_1, \ldots, a_n) = \mathcal{T}_O \quad \text{if and only if} \quad a_1 = \ldots = a_n = 1.$$

Remark 8. *$\mathcal{F}_O$ is composed of contingency tables of X and Y. Indeed, for each $\mathcal{T}_O(a_1, \dots, a_n) \in \mathcal{F}_O$ and for each $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$, the value $\mu_j(x_i) \cdot t_i \cdot a_i$ is a non-negative integer. This is clear because* [10]

$$\mu_j(x_i) \cdot t_i \cdot a_i = \frac{q'_{ij}}{t_i} \cdot t_i \cdot a_i = q'_{ij} \cdot a_i,$$

where $q'_{ij} \in \mathbb{Z}_0^+$ and $a_i \in \mathbb{Z}^+$.

Remark 9. *$\mathcal{F}_O$ is infinite. Observe that all the objects of $\mathcal{F}_O$ can be obtained by varying the choice of $a_1, \dots, a_n$ in $\mathbb{Z}^+$.*

Remark 10. *$\mathcal{T}_O$ defined by Table 18, is the contingency table of $\mathcal{F}_O$ with the smallest cell values. More precisely, let $\mathcal{T}_O(a_1, \dots, a_n) \in \mathcal{F}_O$,*

$$\mu_j(x_i) \cdot t_i \le \mu_j(x_i) \cdot t_i \cdot a_i,$$

for each $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$.

The next proposition exhibits that the contingency tables in $\mathcal{F}_O$ are associated with the same conditional relative frequency distribution [11].

Proposition 4.2. *Let $\mathcal{T}, \tilde{\mathcal{T}} \in \mathcal{F}_O$, then $\mathcal{T}_{Y/X} = \tilde{\mathcal{T}}_{Y/X}$.*

Proof. Let $a_1, \dots, a_n \in \mathbb{Z}^+$ and let $\tilde{a}_1, \dots, \tilde{a}_n \in \mathbb{Z}^+$ so that

$$\mathcal{T} = \mathcal{T}_O(a_1, \dots, a_n) \quad \text{and} \quad \tilde{\mathcal{T}} = \mathcal{T}_O(\tilde{a}_1, \dots, \tilde{a}_n).$$

Hence, given $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$,

$$\mu_j(x_i) \cdot t_i \cdot a_i \quad \text{and} \quad \mu_j(x_i) \cdot t_i \cdot \tilde{a}_i$$

are the cell values in $\mathcal{T}$ and $\tilde{\mathcal{T}}$, respectively.

We denote with b_{ij} and $\tilde{b}_{ij}$ the elements displayed in the i-th row and j-ih column of $\mathcal{T}_{Y/X}$ and $\tilde{\mathcal{T}}_{Y/X}$. Thus, using the standard definition of the relative frequency distribution of Y given X provided in Subsection 2.1, it is easy to verify that

$$b_{ij} = \tilde{b}_{ij} = \frac{\mu_j(x_i)}{\mu_1(x_i) + \ldots + \mu_m(x_i)}. \tag{17}$$

□

Example 4.3. Let $O = \{(\mu_1, \nu_1), (\mu_2, \nu_2), (\mu_3, \nu_3)\}$ be the generalized fuzzy orthopartition of $\{x_1, x_2\}$ defined by Table 17. Then, let us consider the contingency tables $\mathcal{T}_O(1, 1)$ and $\mathcal{T}_O(2, 2)$ of $\mathcal{F}_O$, which respectively coincide with Tables 15 and 20. In line with Proposition 4.2, $\mathcal{T}_O(1, 1)$ and $\mathcal{T}_O(2, 2)$ generate the same relative frequency distribution of Y given X, which is represented by Table 21. In fact, we can immediately check that Eq. (17) holds for each $(i, j) \in \{1, 2\} \times \{1, 2, 3\}$: for example, the entry in the first row and the second column is

$$\frac{\mu_2(x_1)}{\mu_1(x_1) + \mu_2(x_1) + \mu_3(x_1)} = \frac{\frac{5}{14}}{\frac{1}{14} + \frac{5}{14} + \frac{1}{7}} = \frac{5}{8}.$$

[10] Recall that $\mu_j(x_i) = \frac{q'_{ij}}{t_i}$ from (15).

[11] According to the notation introduced in Subsection 2.1, we use the symbols $\mathcal{T}_{Y/X}$ and $\tilde{\mathcal{T}}_{Y/X}$ to denote the relative frequency distributions of Y given X, which respectively derive from the contingency tables $\mathcal{T}$ and $\tilde{\mathcal{T}}$.

Table 21. Relative Frequency Distribution of $Y = \{y_1, y_2, y_3\}$ given $X = \{x_1, x_2\}$ assigned to $\mathcal{T}_O(1, 1)$ and $\mathcal{T}_O(2, 2)$ (Example 4.3).

$\mathcal{T}_{Y/X}$	y_1	y_2	y_3
x_1	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{1}{4}$
x_2	$\frac{4}{9}$	$\frac{1}{9}$	$\frac{4}{9}$

Summing up the results shown until now by this section, a generalized fuzzy orthopartition satisfying Eq. (10) for all pairs of indices, determines infinitely many contingency tables having the same conditional relative frequency distribution.

From now on, for convenience, we indicate with $\mathcal{T}_{Y/X}$ the relative frequency distribution of Y given X, which is related to the contingency tables in $\mathcal{F}_O$.

Remark 11. *From an algorithmic standpoint, the computational cost of constructing $\mathcal{T}_O$ from a given generalized fuzzy orthopartition O depends on Steps 1, 2, and 3, which are repeated for each $i \in \{1, \dots, n\}$, as presented at the beginning of Section 4.*

Step 1: *We need to convert the values $\mu_1(x_i), \dots, \mu_m(x_i)$, and h_i into fractions and reduce them to their lowest terms. This requires $O(m)$ time.*
Step 2: *We compute the least common multiple (l.c.m.) of $p_{i1}, \dots, p_{im}, r_i$, which also takes $O(m)$ time.*
Step 3: *We multiply each $\mu_j(x_i)$ by t_i for every $j \in \{1, \dots, m\}$, which again is done in $O(m)$ time.*

Since these three steps are repeated for each row $i = 1, \dots, n$, the total time complexity for constructing $\mathcal{T}_O$ is $O(n \cdot m)$.

The set $\mathcal{F}_O$ is composed of infinitely many contingency tables. Indeed, one table is determined for each choice of $(a_1, \dots, a_n) \in \mathbb{Z}^+$. However, we may restrict to a finite subset of $\mathcal{F}_O$ by introducing additional constraints. For example, if we require $a_i \leq M$ for each $i \in \{1, \dots, n\}$, then the complexity for finding such particular tables in $\mathcal{F}_O$ becomes $O(n \cdot m \cdot M^n)$.

Definition 4.4. Given $\mathcal{T}_O(a_1, \dots, a_n) \in \mathcal{F}_O$, we denote with $\mathcal{T}'_{Y/X}(a_1, \dots, a_n)$ the relative frequency distribution of Y given X with $\alpha_1, \dots, \alpha_n$ undefined observations deriving from $\mathcal{T}_O(a_1, \dots, a_n)$.

In the next theorem, we determine the values of $\alpha_1, \dots, \alpha_n$ so that O can be obtained from $\mathcal{T}'_{Y/X}(a_1, \dots, a_n)$ by means of Definition 3.3 [12]. Recall that in this case, we say that O and $\mathcal{T}'_{Y/X}(a_1, \dots, a_n)$ are equivalent, since they have the same consistent distributions.

Theorem 4.5. *O is equivalent to $\mathcal{T}'_{Y/X}(a_1, \dots, a_n)$, where $\alpha_1 = s'_1 \cdot a_1, \dots, \alpha_n = s'_n \cdot a_n$.*

Proof. Let $\mathcal{T}_O(a_1, \dots, a_n) \in \mathcal{F}_O$ and let $\mathcal{T}'_{Y/X}(a_1, \dots, a_n)$ be its conditional relative frequency distribution with $\alpha_1, \dots, \alpha_n$ undefined observations, where $\alpha_1 = s'_1 \cdot a_1, \dots, \alpha_n = s'_n \cdot a_n$. Recall that the element of $\mathcal{T}_O(a_1, \dots, a_n)$ are defined by Table 19. Hence, according to Definition 3.3, we need to prove that the equalities

$$\mu_j(x_i) = \frac{a_{ij}}{t_{x_i} + \alpha_i} \quad \text{and} \quad \nu_j(x_i) = 1 - \frac{a_{ij} + \alpha_i}{t_{x_i} + \alpha_i}$$

[12] In Theorem 4.5, we deal with $s'_1, \dots, s'_n$, which are non-negative integers associated with O by (15).

are true for each $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$, where
$a_{ij} = \mu_j(x_i) \cdot t_i \cdot a_i$ (namely, a_{ij} is the value displayed by the i-row and j-column of $\mathcal{T}_O(a_1, \ldots, a_n)$),
$t_{x_i} = \sum_{j=1}^{m} \mu_j(x_i) \cdot t_i \cdot a_i$ (namely, t_{x_i} is the total number of observations for x_i in $\mathcal{T}_O(a_1, \ldots, a_n)$),
$\alpha_i = s'_i \cdot a_i$.

Firstly, let us show that $t_{x_i} = (t_i - s'_i) \cdot a_i$. Indeed,

$$t_{x_i} = \mu_1(x_i) \cdot t_i \cdot a_i + \ldots + \mu_j(x_i) \cdot t_i \cdot a_i = (\mu_1(x_i) + \ldots + \mu_m(x_i)) \cdot t_i \cdot a_i,$$

where $\mu_1(x_i) + \ldots + \mu_m(x_i) = 1 - h_j(x_i)$ by Proposition 3.19. Thus, considering that $h_j(x_i) = \frac{s'_i}{t_i}$ by (15) (recall that we put $h_i = h_j(x_j)$), it is true that

$$t_{x_i} = (1 - h_j(x_i)) \cdot t_i \cdot a_i = \left(1 - \frac{s'_i}{t_i}\right) \cdot t_i \cdot a_i = \left(\frac{t_i - s'_i}{t_i}\right) \cdot t_i \cdot a_i = (t_i - s'_i) \cdot a_i.$$

Hence,

$$\frac{a_{ij}}{t_{x_i} + \alpha_i} = \frac{\mu_j(x_i) \cdot t_i \cdot a_i}{(t_i - s'_i) \cdot a_i + s'_i \cdot a_i},$$

which obviously coincides with $\mu_j(x_i)$. Moreover,

$$1 - \frac{a_{ij} + \alpha_i}{t_{x_i} + \alpha_i} = 1 - \frac{\mu_j(x_i) \cdot t_i \cdot a_i + s'_i \cdot a_i}{(t_i - s'_i) \cdot a_i + s'_i \cdot a_i} = 1 - \left(\frac{\mu_j(x_i) \cdot t_i \cdot a_i}{t_i \cdot a_i} + \frac{s'_i \cdot a_i}{t_i \cdot a_i}\right) =$$
$$1 - \left(\mu_j(x_i) + \frac{s'_i}{t_i}\right) = 1 - (\mu_j(x_i) + h_j(x_i)),$$

Additionally, $1 - (\mu_j(x_i) + h_j(x_i)) = \nu_j(x_i)$ by (2). Finally, we can conclude that $1 - \frac{a_{ij} + \alpha_i}{t_{x_i} + \alpha_i} = \nu_j(x_i)$. □

Therefore, according to Theorem 4.5, O is equivalent to infinitely many conditional relative frequency distributions with undefined observations, one for each contingency table $\mathcal{T}_O(a_1, \ldots, a_n)$ of the class $\mathcal{F}_O$.

Example 4.6. Let $O = \{(\mu_1, \nu_1), (\mu_2, \nu_2), (\mu_3, \nu_3)\}$ be the generalized fuzzy orthopartition of $\{x_1, x_2\}$ represented by Table 17 and let $\mathcal{T}_O(2, 2) \in \mathcal{F}_O$ that is defined by Table 20 (hence, $a_1 = a_2 = 2$). As shown by Example 4.1, $s'_1 = 6$ and $s'_2 = 7$ (this is because $h_1 = 6/14$ and $h_2 = 7/16$). Therefore, let us consider $\mathcal{T}'_{Y/X}$ that derives from $\mathcal{T}_O(2, 2)$ by adding α_1 and α_2 undefined observations, where

$$\alpha_1 = s'_1 \cdot a_1 = 6 \cdot 2 = 12 \quad \text{and} \quad \alpha_2 = s'_2 \cdot a_2 = 7 \cdot 2 = 14.$$

$\mathcal{T}'_{Y/X}$ is defined by Table 22. It is easy to verify that starting from $\mathcal{T}'_{Y/X}$ and applying Definition 3.3, we obtain the elements of O. For example, taking into account $i = 1$ and $j = 2$, we have

$$\mu_2(x_1) = \frac{10}{28} = \frac{5}{14} \quad \text{and} \quad \nu_2(x_1) = 1 - \frac{10 + 12}{28} = 1 - \frac{22}{28} = \frac{6}{28} = \frac{3}{14}.$$

Furthermore, referring again to Theorem 4.5, O is also equivalent to the conditional relative frequency distribution with undefined observations deriving from $\mathcal{T}_O(1, 1)$ (that is $a_1 = a_2 = 1$), where

$$\alpha_1 = s'_1 \cdot a_1 = 6 \cdot 1 = 6 \quad \text{and} \quad \alpha_2 = s'_2 \cdot a_2 = 7 \cdot 1 = 7.$$

This is represented by Table 16. Other distributions that are equivalent to O can be obtained by varying a_1 and a_2 in $\mathbb{Z}^+$, in order to obtain $\mathcal{T}_O(a_1, a_2)$ and $\alpha_1 = s'_1 \cdot a_1$ and $\alpha'_2 = s'_2 \cdot a_2$.

Table 22. $\mathcal{T}'_{Y/X}$ deriving from $\mathcal{T}_O(2,2)$ with $\alpha_1 = 12$ and $\alpha_2 = 14$ (Example 4.6).

$\mathcal{T}'_{Y/X}$	y_1	y_2	y_3
x_1	$\frac{2+k_{11}}{28}$	$\frac{10+k_{12}}{28}$	$\frac{4+k_{13}}{28}$
x_2	$\frac{8+k_{21}}{32}$	$\frac{2+k_{22}}{32}$	$\frac{8+k_{23}}{32}$

The previous theorem states that O is equivalent to each distribution that derives from a table $\mathcal{T}_O(a_1, \dots, a_n)$ of $\mathcal{F}_O$, when $\alpha_1, \dots, \alpha_n$ assume specific values depending from O and $a_1, \dots, a_n$. Vice-versa, we can prove that if a distribution $\mathcal{T}'_{Y/X}$ is equivalent to O, then it must derive from a contingency table of $\mathcal{F}_O$. Thus, the following theorem shows that starting from $\mathcal{F}_O$, we can generate all the relative frequency distributions with undefined observations equivalent to O.

Theorem 4.7. *Let $\mathcal{T}'_{Y/X}$ be a conditional relative frequency distribution with undefined observations that is equivalent to O. Then, $\mathcal{T}'_{Y/X}$ derives from $\mathcal{T} \in \mathcal{F}_O$.*

Proof. We denote with a_{ij} the entry of the i-th row and j-th column of $\mathcal{T}$, where $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$. We have to prove that $\mathcal{T} = \mathcal{T}_O(a_1, \dots, a_n)$, where $a_1, \dots, a_n \in \mathbb{Z}^+$.

Let $i \in \{1, \dots, n\}$, by Definition 3.3 and by Proposition 3.17, since $\mathcal{T}'_{Y/X}$ is equivalent to O,

$$\mu_j(x_i) = \frac{a_{ij}}{t_{x_i} + \alpha_i} \quad \text{and} \quad h_j(x_i) = \frac{\alpha_i}{t_{x_i} + \alpha_i}, \tag{18}$$

for each $j \in \{1, \dots, m\}$.

Furthermore, by (14),

- $\mu_1(x_i), \dots, \mu_m(x_i)$ respectively equal the irreducible fractions $\frac{q_{i1}}{p_{i1}}, \dots, \frac{q_{im}}{p_{im}}$;
- h_i equals the irreducible fraction $\frac{s_i}{r_i}$.

Hence, $t_{x_i} + \alpha_i$ must be a multiple of $p_{i1}, \dots, p_{im}$ and r_i. As a consequence, $t_{x_i} + \alpha_i$ needs to be a multiple of t_i too (recall that t_i is the lowest common multiple of $p_{i1}, \dots, p_{im}$ and r_i). Thus, we can be sure that there exists $a_i \in \mathbb{Z}^+$ such that $t_{x_i} + \alpha_i = t_i \cdot a_i$. Moreover, since $a_{ij} = \mu_j(x_i) \cdot (t_{x_i} + \alpha_i)$ by (18), we get $a_{ij} = \mu_j(x_i) \cdot t_i \cdot a_i$. Hence, $\mathcal{T} = \mathcal{T}_O(a_1, \dots, a_n)$ (recall that $\mathcal{T}_O(a_1, \dots, a_n)$ is defined by Table 19). Ultimately, according to the definition of $\mathcal{F}_O$ given by (16), we can conclude that $\mathcal{T} \in \mathcal{F}_O$. □

Example 4.8. Let us consider $\mathcal{T}'_{Y/X}$ defined by Table 22. As explained in Example 4.6, $\mathcal{T}'_{Y/X}$ is equivalent to $O_{Y/X}$ represented by Table 17. In addition, $\mathcal{T}'_{Y/X}$ derives from Table 20, which coincides with $\mathcal{T}_O(2, 2)$. Lastly, as shown by Example 4.1, $\mathcal{T}_O(2, 2) \in \mathcal{F}_O$.

The main Section 4 results presented up to this point are summarized in Figure 7 [13].

The next theorem provides a necessary and sufficient condition for a generalized fuzzy orthopartition to be equivalent to a conditional relative frequency distribution with undefined observations.

[13]Recall that a conditional relative frequency distribution with undefined observations and a generalized fuzzy orthopartition are equivalent when they represent the same class of conditional relative frequency distributions.

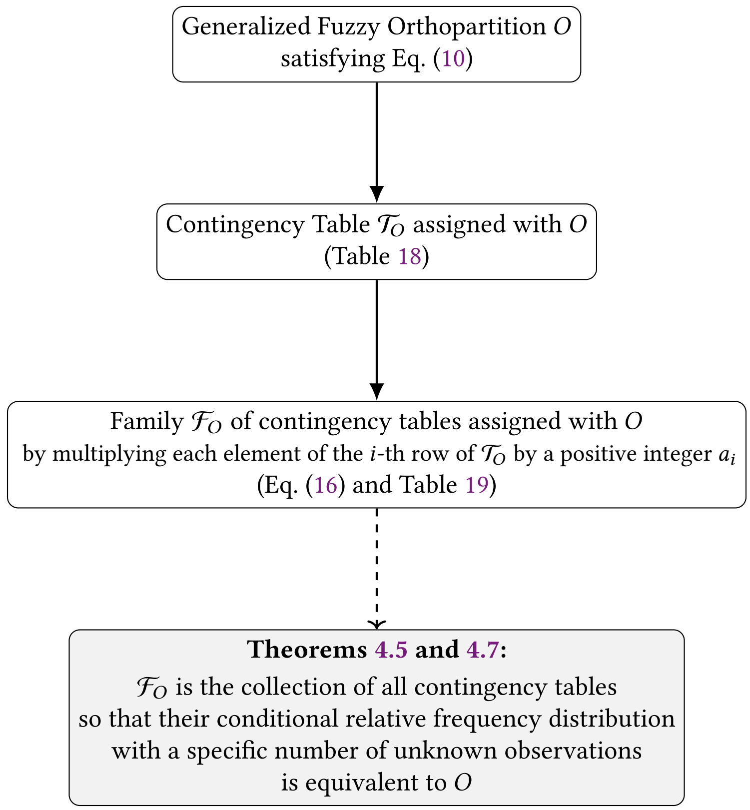

Fig. 7. Logical structure of Section 4 results (up to Theorem 5): from Generalized Fuzzy Orthopartitions to Conditional Relative Frequency Distributions with Undefined Observations.

Theorem 4.9. *Let $O = \{(\mu_1, \nu_1), \ldots, (\mu_m, \nu_m)\}$ be a generalized fuzzy orthopartition of $\{x_1, \ldots, x_n\}$. Then, O is equivalent to a conditional relative frequency distribution with undefined observations if and only if Eq.* (10) *holds for each $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$.*

Proof. **(⇒).** This implication follows from Proposition 3.15.

(⇐). This implication follows from Theorem 4.5. Indeed, when Eq. (10) is true for each $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$, we can consider the collection of contingency tables $\mathcal{F}_O$ defined by (16). Then, from each element in $\mathcal{F}_O$, we can obtain a conditional relative frequency distribution with undefined observations that is equivalent to O.

□

Thus, Theorem 4.9 identifies a subclass of generalized fuzzy orthopartitions that can be viewed as conditional relative frequency distributions with undefined observations. Such a relationship is schematized in the Euler-Venn diagram of Figure 1.

In Section 3, given a conditional relative frequency distribution $\mathcal{T}_{Y/X}$ and $\alpha_1, \ldots, \alpha_n$ undefined observations, we have constructed $\mathcal{T}'_{Y/X}$ and proved that it is equivalent to $O_{Y/X}$. Vice versa, according to this section, we can find out a family of conditional relative frequency distributions with undefined observations that are equivalent to $O_{Y/X}$. Finally, as stated by the following theorem, given $O_{Y/X}$, we can obtain $\mathcal{T}'_{Y/X}$ from $\mathcal{F}_{O_{Y/X}}$ when $a_1, \ldots, a_n$ assume specific values.

Theorem 4.10. *Let $\mathcal{T}'_{Y/X}$ be a conditional relative frequency distribution with $\alpha_1, \ldots, \alpha_n$ undefined observations. Then, $\mathcal{T}'_{Y/X}$ derives from $\mathcal{T}_{O_{Y/X}}(a_1, \ldots, a_n)$, where $a_i = \dfrac{t_{x_i} + \alpha_i}{t_i}$ for each $i \in \{1, \ldots, n\}$* [14].

Proof. Let $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$, recall that $\dfrac{a_{ij} + k_{ij}}{t_{x_i} + \alpha_i}$ indicates the entry of the i-th row and j-th column of $\mathcal{T}'_{Y/X}$ (see Table 7). By Table 19, $\mu_j(x_i) \cdot t_i \cdot a_i$ is the entry of the i-th row and j-th column of $\mathcal{T}_{O_{Y/X}}(a_1, \ldots, a_n)$. Then, we can immediately verify that

$$a_{ij} = \mu_j(x_i) \cdot t_i \cdot a_i, \quad \text{where } a_i = \frac{t_{x_i} + \alpha_i}{t_i}.$$

This is clear because $\mu_j(x_i) = \dfrac{a_{ij}}{t_{x_i} + \alpha_i}$, by considering that the elements of $O_{Y/X}$ are given by Definition 3.3. □

Example 4.11. Let us consider $\mathcal{T}'_{Y/X}$ defined by Table 22. As discussed in the previous examples, $\mathcal{T}'_{Y/X}$ is equivalent to $O_{Y/X}$ represented by Table 17. Namely, $O_{Y/X}$ can be obtained from $\mathcal{T}'_{Y/X}$, applying Definition 3.3. Moreover, we have previously verified that $\mathcal{T}'_{Y/X}$ derives $\mathcal{T}_O(a_1, a_2)$ (Table 20) with $a_1 = a_2 = 2$. Thus, in line with Theorem 4.10, $a_1 = \dfrac{t_{x_1} + \alpha_1}{t_1} = \dfrac{28}{14} = 2$ and $a_2 = \dfrac{t_{x_2} + \alpha_2}{t_2} = \dfrac{32}{16} = 2$.

The links between the results provided in Sections 3 and 4 are highlighted in Figure 8.

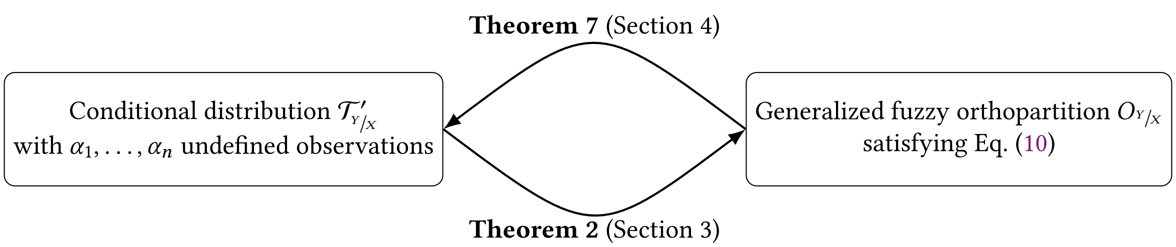

Fig. 8. Connection between Sections 3 and 4.

As for the results exhibited in Section 3, the previous correspondence between generalized fuzzy orthopartitions and conditional relative frequency distributions with undefined observations can be reformulated by exchanging the roles of X and Y. In this dual case, given a generalized fuzzy orthopartition $O = \{(\mu_1, \nu_1), \ldots, (\mu_n, \nu_n)\}$ of $\{y_1, \ldots, y_m\}$, we can generate a class of relative frequency distributions of X given Y with undefined observations.

Illustrative Case Studies from Multiple Domains

To strengthen the empirical assessment of the proposed approach, in addition to the illustrative example on tourism already proposed at the end of Section 3, we mention four realistic application scenarios from different domains, where data uncertainty is common and generalized fuzzy orthopartitions become useful models.

[14] Let us underline that $\dfrac{t_{x_i} + \alpha_i}{t_i}$ is a positive integer, since $t_{x_i} + \alpha_i$ is a multiple of t_i, as discussed in the proof of Theorem 4.7.

(a) A first example concerns the frequency of graduates across fields of study and universities in a given country. For each university (variable X), we compute the distribution of graduates among fields (variable Y) as shown in Figure 9 (a). Moreover, we simulate uncertainty by adding a fixed number of unassigned records per university (for example, considering graduates who have obtained a degree in a field that is only partially or not directly related to the required areas of study). This newly derived distribution is represented by a generalized fuzzy orthopartition, highlighting its ability to compactly capture families of uncertain frequency tables and to support comparison and semantic interpretation across different institutions.

(b) In the healthcare domain, generalized fuzzy orthopartitions can be used to represent uncertainty in patient outcomes across treatment categories. For each treatment type (variable X), we compute the distribution of patient outcomes (variable Y), such as full recovery, partial improvement, or no change (see Figure 9 (b)). In real-world settings, these outcomes may be partially recorded or subject to observational noise. This may occur, for instance, when patient follow-up data are missing, outcome categories are ambiguously defined, or information is self-reported rather than clinically assessed. A fuzzy orthopartition allows encoding such uncertainty by expressing each treatment-outcome distribution through degrees that reflect plausible frequency intervals, thereby enabling comparative analyses under imperfect data conditions.

(c) In the financial sector, the framework can be applied to model the uncertain distribution of customer behaviors across market segments. Here, market segments or customer profiles (variable X) are associated with behavioral categories (variable Y), such as loan default, investment type, or spending pattern (see Figure 9 (c)). Due to missing or anonymized transaction data, the observed frequencies may be incomplete or noisy. Such uncertainty may arise from limitations in data collection systems, privacy-preserving mechanisms that mask or aggregate transactions, or inconsistencies in how financial behaviors are categorized across institutions. Using a fuzzy orthopartition to represent these distributions enables reasoning about client behavior under uncertainty, thus supporting risk profiling, market segmentation, and data-driven financial strategies.

(d) In the context of secondary education, generalized fuzzy orthopartitions can be used to represent uncertainty in the classification of student performance across subject areas. For each school (variable X), one can compute the distribution of students who attain proficiency skills in diverse areas (variable Y) as shown in Figure 9 (d). However, due to factors such as incomplete test participation, accommodations for special needs, or inconsistencies in grading criteria, the actual frequencies may not be fully reliable. Fuzzy orthopartitions can model this uncertainty by assigning partial membership to performance levels, enabling more nuanced analysis of educational outcomes and equity across institutions.

These examples demonstrate the versatility of generalized fuzzy orthopartitions in handling structured categorical data with various forms of uncertainty, and their potential for integration into domain-specific decision-making processes.

5 Conclusions and Future Directions

The present work explores generalized fuzzy orthopartitions within the context of descriptive statistics. Indeed, a specific class of generalized fuzzy orthopartitions can be understood as models for describing conditional relative frequency distributions characterized by uncertainty.

Let us list a few directions for extending our results. Firstly, we will propose integrating generalized fuzzy orthopartitions with the *theory of evaluative linguistic expressions* [32], in order to contribute to *eXplainable Artificial Intelligence (XAI)* [33] by linking numerical uncertainty representations to qualitative and interpretable descriptions. Such integration would allow interval-valued information—captured by generalized fuzzy orthopartitions—to be mapped onto linguistic terms (e.g., “likely,” “moderate,” “uncertain”), thereby enabling a transition from

Graduates	Field 1	...	Field m
University 1	g_{11}	...	g_{1m}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
University n	g_{n1}	...	g_{nm}

(a) Contingency Table of X =*"University"* and Y = *"Field"*.

Patients	Outcome 1	...	Outcome m
Treatment 1	p_{11}	...	p_{1m}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
Treatment n	p_{n1}	...	p_{nm}

(b) Contingency Table of X =*"Treatment"* and Y = *"Outcome"*.

Customers	Behavior 1	...	Behavior m
Segment 1	c_{11}	...	c_{1m}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
Segment n	c_{n1}	...	c_{nm}

(c) Contingency Table of X =*"Segment"* and Y = *"Behavior"*.

Students	Area 1	...	Area m
School 1	s_{11}	...	s_{1m}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
School n	s_{n1}	...	s_{nm}

(d) Contingency Table of X =*"School"* and Y = *"Area"*.

Fig. 9. Case Studies.

numerical tables to natural language explanations. This connection opens a pathway to symbolic interpretability in AI systems and aligns with recent efforts to construct linguistic interpretations of uncertainty in structured data models [34]. Then, we will investigate the relationships between generalized fuzzy orthopartitions and *fuzzy rough sets* [35], to explore possible theoretical connections and their potential applications to approximate reasoning and granular classification. As a next step, we intend to study the correspondence between generalized fuzzy orthopartitions and conditional relative frequencies appearing as intervals. Such a type of uncertainty could arise, for example, in the presence of the so-called *approximators* that are linguistic hedges such as "about" or "more than" as explained in [36]. Moreover, we want to investigate generalized fuzzy orthopartitions in further uncertainty frameworks, taking as a reference the models studied in [37]. In addition, we plan to use generalized fuzzy orthopartitions to extend machine learning techniques based on conditional frequency distributions like *Log-Linear Models* [38], *Conditional Random Fields (CRFs)* [39], and *Markov Chains* [40], when not all data can be observed. Fuzzy orthopartitions can represent classifications in which object membership is inherently partial and uncertain, providing a valuable framework for refining the performance of Large Language Models (LLMs). For instance, in the context of the *HeReFaNMi project – Health-Related Fake News Mitigation* – a language model is used to generate synthetic fake news by manipulating factual information. These generated examples are organized into three levels of depth: easy, medium, and hard [41, 42]. This artificial dataset is then employed to train models aimed at detecting fake news. By extending the discrete depth classification into a fuzzy one, we can allow each fake news item to partially belong to different levels, reflecting the uncertainty and overlap inherent in real-world misinformation. Such a fuzzy representation could lead to more nuanced training data and, ultimately, to improved detection accuracy. Another important direction for future work concerns the practical implementation of the proposed framework. While we have provided a theoretical analysis of the computational complexity for constructing generalized fuzzy orthopartitions, further investigation is needed to assess the scalability of the method when applied to large datasets with high-dimensional categorical variables. Future research could explore heuristic strategies, sampling techniques, or constraint-based formulations to efficiently approximate representative distributions within feasible computational bounds. Moreover, we plan to design and release a prototype software tool—possibly as an R or Python package—that automates the construction and analysis of generalized fuzzy orthopartitions from empirical data. This would support broader adoption of the framework and facilitate its integration into pipelines for interpretable machine learning and decision support systems.

Acknowledgments

This work was partially supported by the MUR under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca

References

[1] H Witten Ian and Frank Eibe. Data mining: Practical machine learning tools and techniques, 2005.
[2] Max Kuhn and Kjell Johnson. *Feature engineering and selection: A practical approach for predictive models*. Chapman and Hall/CRC, 2019.
[3] Samina Khalid, Tehmina Khalil, and Shamila Nasreen. A survey of feature selection and feature extraction techniques in machine learning. In *2014 science and information conference*, pages 372–378. IEEE, 2014.
[4] Max Kuhn, Kjell Johnson, et al. *Applied predictive modeling*, volume 26. Springer, 2013.
[5] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. *An introduction to statistical learning*, volume 112. Springer, 2013.
[6] A Rogier T Donders, Geert JMG Van Der Heijden, Theo Stijnen, and Karel GM Moons. A gentle introduction to imputation of missing values. *Journal of clinical epidemiology*, 59(10):1087–1091, 2006.
[7] Joseph L Schafer. Multiple imputation: a primer. *Statistical methods in medical research*, 8(1):3–15, 1999.
[8] Scott R Eliason. *Maximum likelihood estimation: Logic and practice*, volume 96. Sage, 1993.
[9] Bradley P Carlin and Thomas A Louis. *Bayesian methods for data analysis*. CRC press, 2008.
[10] Steven Abney. Bootstrapping. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 360–367, 2002.
[11] Sarah Gold, Edward White, William Roeder, Mike McAleenan, Christine Schubert Kabban, and Darryl Ahner. Probabilistic contingency tables: an improvement to verify probability forecasts. *Weather and Forecasting*, 35(2):609–621, 2020.
[12] SM Taheri, G Hesamian, and R Viertl. Contingency tables with fuzzy information. *Communications in Statistics-Theory and Methods*, 45(20):5906–5917, 2016.
[13] Hisao Miyano and Eric J Beh. Interval-valued scaling of successive categories. *Advanced Studies in Behaviormetrics and Data Science: Essays in Honor of Akinori Okada*, volume 5:197–209, 2020.
[14] Ozgur Yanmaz, Cigdem Kadaifci, and Erhan Bozdag. A new approach to correspondence analysis based on interval-valued hesitant fuzzy sets. *International Journal of Information Technology & Decision Making*, 21(06):1749–1776, 2022.
[15] SQ Shan, YB Bai, and XJ Wang. Water quality assessment based on interval-valued data cluster analysis. *Desalin Water Treat*, 213:84–90, 2021.
[16] DS Hooda and MS Barak. Estimation of missing values in fuzzy matrices (fm) and interval-valued fuzzy matrices (ivfm). *Life Cycle Reliability and Safety Engineering*, 9:241–245, 2020.
[17] Stefania Boffa and Davide Ciucci. Unifying credal partitions and fuzzy orthopartitions. *Information Sciences*, 674:120725, 2024.
[18] Davide Ciucci, Stefania Boffa, and Andrea Campagner. Orthopartitions in knowledge representation and machine learning. In *International Joint Conference on Rough Sets*, pages 3–18. Springer, 2022.
[19] Enrique H Ruspini. A new approach to clustering. *Information and control*, 15(1):22–32, 1969.
[20] Andrea Campagner and Davide Ciucci. Orthopartitions and soft clustering: soft mutual information measures for clustering validation. *Knowledge-Based Systems*, 180:51–61, 2019.
[21] Lotfi A Zadeh et al. Fuzzy sets, information and control. *Information and control*, 8(3):338–353, 1965.
[22] Stefania Boffa and Davide Ciucci. Logical entropy and aggregation of fuzzy orthopartitions. *Fuzzy Sets and Systems*, 455:77–101, 2023.
[23] Stefania Boffa and Davide Ciucci. Orthopartitions and possibility distributions. *Fuzzy Sets and Systems*, 455:53–76, 2023.

[24] Stefania Boffa, Andrea Campagner, and Davide Ciucci. Partially-defined equivalence relations: Relationship with orthopartitions and connection to rough sets. *Information Sciences*, 657:119941, 2024.
[25] Thierry Denœux and M-H Masson. Evclus: evidential clustering of proximity data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):95–109, 2004.
[26] Stefania Boffa, Davide Ciucci, and Christophe Marsala. Extending intuitionistic operations, orderings, and entropy measures on generalized fuzzy orthopartitions. *Fuzzy Sets and Systems*, 513:109381, 2025.
[27] Xibei Yang, Tsau Young Lin, Jingyu Yang, Yan Li, and Dongjun Yu. Combination of interval-valued fuzzy set and soft set. *Computers & Mathematics with Applications*, 58(3):521–527, 2009.
[28] Jan M Zytkow. Granularity refined by knowledge: contingency tables and rough sets as tools of discovery. In *Data Mining and Knowledge Discovery: Theory, Tools, and Technology II*, volume 4057, pages 82–91. SPIE, 2000.
[29] Lynne Billard and Edwin Diday. *Symbolic data analysis: Conceptual statistics and data mining*. John Wiley & Sons, 2012.
[30] Alan Agresti et al. An introduction to categorical data analysis, 1996.
[31] Krassimir Atanassov. Intuitionistic fuzzy sets. In *Intuitionistic fuzzy sets*, pages 1–137. Springer, 1999.
[32] Vilém Novák. A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy sets and systems*, 159(22):2939–2969, 2008.
[33] Dang Minh, H Xiang Wang, Y Fen Li, and Tan N Nguyen. Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, :1–66, 2022.
[34] Stefania Boffa and Davide Ciucci. Three-way decisions with evaluative linguistic expressions. *International Journal of Approximate Reasoning*, 164:109080, 2024.
[35] Didier Dubois and Henri Prade. Rough fuzzy sets and fuzzy rough sets. *International Journal of General System*, 17(2-3):191–209, 1990.
[36] Scott Ferson, Jason O'Rawe, Andrei Antonenko, Jack Siegrist, James Mickley, Christian C Luhmann, Kari Sentz, and Adam M Finkel. Natural language of uncertainty: numeric hedge words. *International Journal of Approximate Reasoning*, 57:19–39, 2015.
[37] Fabio Cuzzolin. Uncertainty measures: A critical survey. *Information Fusion*, Volume 114:102609, 2024.
[38] Ronald Christensen. *Log-linear models and logistic regression*. Springer Science & Business Media, 2006.
[39] Charles Sutton, Andrew McCallum, et al. An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, 4(4):267–373, 2012.
[40] James R Norris. *Markov chains*. Number 2. Cambridge university press, 1998.
[41] Walid Taib, Idriss Saadallah, Abdelali Ichou, Abderrahmene Hamdi, Alessandro Bruno, Pier Luigi Mazzeo, Aladine Chetouani, Marouane Tliba, and Mohamed Amine Kerkouri. Health misinformation detection: A chunking strategy integrated to retrieval-augmented generation. In *2nd Workshop on Artificial Intelligence for Perception and Artificial Consciousness, AIxPAC 2024; Bolzano; Italy; 28 November 2024 through; Code 206893*, volume 3923, pages 41–46. CEUR Workshop Proceedings, 2025.
[42] Alessandro Bruno, Pier Luigi Mazzeo, Aladine Chetouani, Marouane Tliba, and Mohamed Amine Kerkouri. Insights into classifying and mitigating llms' hallucinations. *arXiv preprint arXiv:2311.08117*, , 2023.

Received 3 January 2025; revised 27 May 2025; accepted 2 June 2025