

Article

Not peer-reviewed version

---

# Weighted Entropy: Evaluating Feasible Weights with an Inversion Procedure and Eliciting a New Probability-Possibility Transformation

---

[José Pinto Casquilho](#)\* and [Vitor Matos](#)\*

Posted Date: 17 December 2024

doi: 10.20944/preprints202412.1479.v1

Keywords: weighted entropy; discrete setting; utilities; maximum point; inversion procedure; feasible weights; fuzzy weighted entropy; probability-possibility transformation; consistency



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

## Article

# Weighted Entropy: Evaluating Feasible Weights with an Inversion Procedure and Eliciting a New Probability-Possibility Transformation

José Pinto Casquilho\*<sup>1</sup> and Vítor Matos\*<sup>2</sup>

<sup>1</sup> Postgraduate and Research Program, Universidade Nacional Timor Lorosa'e, Dili, Timor-Leste

<sup>2</sup> Centro de Matemática, Faculty of Economics, Universidade do Porto, Porto, Portugal

\* Correspondence: jose.casquilho@untl.edu.tl (J.P.C.); vmatos@fep.up.pt (V.M.)

**Abstract:** In this article, after an extended literature survey on the topics of weighted entropy and possibility theory, including probability-possibility transformations, we present some new results regarding the analytical study of the maximum point of weighted entropy. Then, we build an inversion procedure which allows for computing feasible weights given an optimal point solution, which shows to be insensitive to a positive linear scaling. From there, we associate the calculated feasible weights with a possibility distribution and show that the inversion procedure can be interpreted to elicit a new probability-possibility transformation, which is studied from the perspective of a set of axioms including consistency and preference order preservation. Numeric examples are outlined and some related criteria are evaluated while the new probability-possibility transformation is compared with other standard possibility distributions mentioned in the literature, with the results showing an admissible performance. However, there is an intrinsic limitation regarding a least upper bound of the optimal point of weighted entropy and another restriction concerning a threshold for consistency, but an alternative is still mentioned that can be considered for future work related to this subject.

**Keywords:** weighted entropy; discrete setting; utilities; maximum point; inversion procedure; feasible weights; fuzzy weighted entropy; probability-possibility transformation; consistency

**MSC:** 94A17; 62E86; 90C30

## 1. Introduction

Weighted Shannon entropy was initially published in 1968 by Marianne Belis and Silviu Guiaşu, characterized within the scope of a quantitative-qualitative measure of information in cybernetic systems [1], introducing the concept of utility of an event associated to goals. To avoid ambiguity, in this paper we will be dealing with the quantity denoted by Equation (1), regarding a discrete and finite setting of events.

$$H_w = - \sum_{i=1}^n w_i p_i \log p_i \quad (1)$$

Equation (1) is relative to a complete probability distribution  $P = (p_1, p_2, \dots, p_n)$  with domain in the simplex  $\Delta^{n-1} = \{(p_1, \dots, p_n) : p_i \geq 0, \sum_{i=1}^n p_i = 1\}$ , and the weights will be considered strictly positive real numbers ( $w_i > 0$ ). Later, we clarify that the case  $w_j = 0$  for some  $j$  shows to be less important for the scope of this article.

Utilities have a long history in the field of conceptualizing goals and preferences in the topics of Moral and Economics (e.g., [2-4]), with utility theory mainly based in the notion of mathematical expectation. A recent review of utility concepts within several theories – since its approach by

Aristotle and discussion by Jeremy Bentham in the last quarter of 18th century – can be found in Monroe *et al.* [5].

Yet, in the very same year (1968) and independently, Lotfi Zadeh presented the same mathematical structure concerning the theme of probability measures of fuzzy events [6], where the weights are replaced by the values of the membership function, denoting another component of uncertainty associated with fuzzy events, a concept he had previously defined as a class of objects with a grade of membership ranging between zero and one [7].

Probability is associated with measuring random uncertainty, while fuzziness – and, more specifically, possibility theory – is considered a setting for handling imprecision due to epistemic uncertainty, referred to as a lack, poverty, or limited amount of information [8], meaning that possibility and probability do not capture the same facets of ignorance, a possibility distribution representing certain but fuzzy evidence (e.g., [9-10]). A possibility distribution can be considered a particular fuzzy set of mutually exclusive possible values ([11], p. 860), and yet, possibility, probability and utility, can be combined in hybrid frameworks under the scope of information fusion (e.g., [12-14]).

This paper is organized as follows: in Section 2 we review the literature background concerning weighted entropy and its developments or applications, mainly based on utilities; next, in another subsection, we address a note on a possibility-weighted entropy and possibility theory including probability-possibility transformations. In Section 3, we present some new results regarding the analytical study of the optimal point of weighted entropy and the inversion procedure is also outlined. In Section 4, the results are interpreted in terms of a probability-possibility transformation. Last, in Section 5, we focus on discussion assessed with a set of axioms and some numerical examples, then proceeding to highlight some limitations and possible further developments on this subject.

## 2. Literature Review

In this section, we will address separately background surveys concerning weighted Shannon entropy associated with finite and discrete settings, and then – linked by the concept of entropy of fuzzy events – a note on possibility theory and probability-possibility transformations. In either case, we follow a broad chronological order in the surveys, but not strictly, as there are exceptions.

### 2.1. On Weighted Entropy

Ralph Hartley introduced a measure of information defined as the logarithm of the number of possible symbol sequences [15], conceived with no reference to the concept of probability but else to possibility in an usual sense, and later named Hartley entropy (e.g., [16]). After two decades, Claude Shannon introduced a measure of uncertainty of the outcome in a random experiment using the logarithm of the number of available choices ([17], p. 7) with the mean value of the random variable called entropy [18], a name that was proposed to him by John von Neumann (e.g., [19]).

In summary, considering a discrete random variable with  $n$  outcomes such as  $X = \{x_1, x_2, \dots, x_n\}$ , each associated with an elementary event with probabilities  $\Pr[X = x_i] = p_i$  for  $i = 1, \dots, n$ , the entropy is evaluated as  $H = -k \sum_{i=1}^n p_i \log p_i$  with  $k > 0$  being a constant that allows for changing the base of the logarithms; for the sake of simplicity we will consider  $k = 1$  when using natural logarithms. It should be noted that when the events are equally probable, Shannon entropy reduces to Hartley entropy.

In fact, the concept, and even the name ‘entropy’, was already in use in statistical physics in works of Boltzmann and Gibbs since the 19<sup>th</sup> century (e.g., [20]), and the consistency between the interpretations in statistical mechanics and information theory was discussed by E. T. Jaynes concluding that uncertainty and information should be considered synonymous [21]. There are other entropies, including weighted versions, such as the framework conceived by Pal and Pal [22], but those will not be considered in this paper.

After Belis and Guiaşu characterization of weighted entropy [1], three years later Silviu Guiaşu carried out an axiomatic and analytical study [23], deriving what he called the principle of maximum information associated to the maximum point and the corresponding maximum value. From then

until today, there have been many developments and applications, and we will mention several, with a focus on discrete settings and pioneer research.

Longo opened a discussion on coding procedures involving qualitative parameters as utilities, embodied in what he named useful self-information whose mean value was called ‘useful entropy’ [4], and Skala referred to it stating that could be conceived as pertaining to the field of semantic information [24]. Also in the 70’s, Bouchon applied the concept of ‘useful information’ to questionnaires, clarifying that the importance should else be ascribed to preference(s), defined as the product of utility and probability associated with an event [25]. Two years later, Sharma *et al.* published results concerning measures on ‘useful information’ [26], while Aggarwal and Picard generalized into information measures with preference(s) [27].

In the beginning of the 80’s, Dial and Taneja provided a characterization of a generalization referred to as weighted entropy of type  $(\alpha, \beta)$  [28] and Kapur made a comparative survey of various measures of entropy including weighted entropy [19] which he showed that could be used to explain the entropy of a Markov chain. The framework was also applied in the field of Geography where it was used to discuss the spatial pattern of aggregation in cities [29], and in Economics too, considering a state-value weighted entropy as a measure of investment risk [30]. Still, Guiaşu resumed the subject using the concept now applied to data clustering and deriving a connection with Sturges’ rule [31], while Jumarie dealt with the observed weighted entropy applied to pattern recognition [32].

Entropy of fuzzy events, considering two kinds of uncertainty – probabilistic and possibilistic – was discussed by Criado and Gachechiladze using weighted Shannon entropy and other frameworks [33]. Also, when discussing fuzzy sets, Cios *et al.* referred to weighted entropy emphasizing that choosing the weights like  $w_i = -p_i/\log p_i$  transforms  $H_w$  into a sum of squared probabilities [34], which is known as Simpson index [35-36]. Still, cost-weighted entropy was used to select measurements with most discernment at the lowest cost in the context of flexible manufacturing systems [37].

In the new millennium, Guiaşu presented and discussed conditional and weighted measures of ecological diversity [38] – therein revisiting weighted entropy – where the weights could reflect supplementary information about the abundance, the economic significance, or the conservation values of the species. Srivastava and Maheshari emphasized a new weighted information generating function, whose derivative at a specific point originates weighted entropy [39], a result which could be considered akin to previous deductions by Hooda and colleagues (e.g., [40]).

Another axiomatic framework for weighted entropies was provided by Ebanks, under the scope of generalizations of the entropies of degree  $\alpha$  with utility values [41], and one can find an application that discusses the composition of the landscape mosaic using a parametric generalization of weighted entropy [42], previously addressed in 1997 [43]. The framework was also applied to security quantification [44].

Moving to the 2020’s, Singer *et al.* built an information-gain ratio measure for selecting the classifying attributes in ordinal classification trees using weighted entropy [45] and, by taking into account specific qualitative characteristics of each event, were formulated measures of directed information [46]. Still, a novel definition was proposed to improve clustering performance for small and diverse datasets considering intra-class and inter-class weighted entropies for categorical and numeric attributes [47]. Under health sciences domain, weighted entropy was used for detecting disease association with genetic rare variants of patients [48], and a weighted entropy method was used to assess how the indicators affected the relative importance of nodes in a network [49].

Weighted entropy, including the entropy of a fuzzy event in the sense of Zadeh, was revisited under the scope of artificial intelligence with a focus on monotonicity [50]. Also, Aggarwal discusses a weighted entropy framework for randomness and fuzziness, pointing out some hermeneutic difficulties arising from the interpretation of the entropy of a fuzzy event as a measure of uncertainty [51] noting that, in the context, it does not provide an intuitive measure [52].

We will delve deeper into this connection although from a different perspective, emphasizing a special case of fuzziness regarding a possibility distribution, reminding that there are two conflicting views on this subject: from one perspective, knowledge representation in possibility theory is driven



by the principle of minimal specificity [8], stating that any hypothesis not known to be impossible cannot be ruled out; however, in another view, when moving from probability to possibility it is claimed that the conversion should satisfy the preference preservation constraint and the criterion of maximal specificity [9], what is particularly emphasized when referring to measurements.

An informational principle usually evoked, similar to the closed-world assumption, says that any situation not yet observed is tentatively considered as impossible ([11], p. 862), which is a correlate of the frequency interpretation of probability that is considered characteristic of the ensemble, and, without the ensemble, cannot be said to exist [53].

Next, we will remind possibility theory *sensu* Zadeh, evoking that probabilistic and possibilistic settings are not two equivalent representations of uncertainty, as the possibilistic representation is weaker because it handles imprecision or incomplete knowledge, and possibility measures are based on an ordering structure rather than an additive one (e.g. [54]).

### 2.1. On Possibility-weighted Entropy and Probability-possibility Transformations

An important theme for this article concerns the recovery of weighted entropy from the original formulation by Zadeh as the entropy of a fuzzy event ([6], p. 426) referred to as Equation (1) when replacing  $w_i$  by  $\mu_i$  (for  $i = 1, \dots, n$ ), concerning the value of the membership function associated with an event expressed with an outcome  $\{x_i\}$ , occurring with probability  $p_i$  and such that  $\mu_i = \mu(x_i)$ , with  $0 \leq \mu_i \leq 1$  measuring the degree of feasibility of  $\{x_i\}$  for other reasons than randomness: for instance, a lack of proper identification or any form of incomplete knowledge about the situation.

It was also Lotfi Zadeh who, in 1978, inaugurated the concept of fuzzy sets as a basis for a theory of possibility [55] – inspired by a paper on possible automata, where the authors addressed the notion of possibility required in the analysis of system stability and reliability [56] – clarifying that, in general, a variable may be associated both with a possibility distribution and a probability distribution. He stated that the possibility distribution function denoted like  $\pi_X(x)$ , meaning the possibility that  $X = x$ , is numerically equal to  $\mu_F(x)$ ,  $F$  being a fuzzy set and  $\mu_F$  its membership function with  $\mu_F(x)$  assessing the grade of compatibility of  $x$  with the concept labeled  $F$ .

Stated briefly, a possibility distribution restricts a set of possible values for a variable of interest in an elastic way and this may be used for representing uncertainty for an ill-known state of the world ([11], p. 859). Or, in another way, a possibility distribution is a mapping to the unit interval, describing what one knows about the more or less plausible values of the uncertain variable and these values are assumed to be mutually exclusive, since the uncertain variable takes on only one value, the true one ([57], p. 30).

For the purposes of this article, we will consider membership values  $\mu_i$  to be possibility values denoted as  $\pi_i = \pi(x_i)$ , and interpretable as the degree of feasibility with respect to the adequate identification of a specific occurrence. As Dubois clarified ([58], p. 47), the connection between possibility theory and probability theory can be fruitful in the scope of statistical reasoning when uncertainty due to variability of observations should be distinguished from uncertainty due to incomplete information.

Zadeh also introduced a (weak) possibility-probability consistency principle associated with a variable  $X$  taking the values  $\{x_1, \dots, x_n\}$  with possibility distribution  $\pi = (\pi_1, \dots, \pi_n)$  and probability distribution  $p = (p_1, \dots, p_n)$ , then defining the degree of consistency of  $p$  with  $\pi$  expressed by the inner product of the corresponding vectors,  $\gamma = \sum_{i=1}^n \pi_i p_i$  and observing that the computation of  $\gamma$  corresponds to the heuristic observation that a lessening of the possibility of an event tends to lessen its probability, but not the opposite. The product  $\gamma$  approaches the *infimum* value zero when the vectors are almost orthogonal while, on the contrary, the maximum value  $\gamma = 1$  occurs when all events with non-zero probability are completely possible ( $\pi_i = 1, \forall i$ ).

In fact, it is said that when representing knowledge,  $\pi(x_i) = 1$  means that  $\{x_i\}$  is certainly possible because this value or state has been actually observed, and in turn  $\pi(x_j) = 0$ , when representing knowledge, means that nothing is known about this value  $\{x_j\}$  which has not been observed [59], what can imply that  $\{x_j\}$  should be ruled out; the limit case  $\pi_j = 0$  means

impossibility, either because of the impossibility of the event at stake – which corresponds to  $p(x_j) = 0$  in the axiomatic definition of probability – or of identifying the specific outcome because of a complete lack of feasibility.

Next, we will focus on reviewing some pioneering conceptualizations on this topic, and then we will move on to recent developments. Some intermediate references will be addressed later in the discussion of results.

Dubois and Prade notably elaborated on this subject, with developments that go on since 1980 until nowadays, for instance soon observing that what is probable is certainly possible and what is inevitable, or necessary, is certainly probable, thus concluding that consistency entails that the degree of possibility of an event must be equal to or greater than its degree of probability ([60], p. 138).

Although a possibility measure was conceived relating to nested sets, it was clarified that considering that  $X$  is a finite set equipped with a probability measure and introducing a point-to-set mapping  $\Gamma$  from  $X$  to some set  $S$ , then a possibility measure is completely specified by the knowledge of its restriction to the set of singletons of  $S$ :  $\{\pi_i = \Pi(\{s_i\}), s_i \in S\}$ , and because  $\Pi(S) = 1$ , there exists an order  $i$  such that  $\pi_i = 1$  (e.g., [54], [61]). A possibility measure is driven by the ‘maxitivity’ property.

Dubois and Prade also introduced a distinction between physical and epistemic possibility, and derived a pioneer probability-possibility transformation from statistical evidence expressed in a histogram [62], which preserves the probability of elementary events, conveying the concept of consistency between possibility and probability distributions, namely  $\pi_{(i)} \geq p_{(i)}, \forall i$ . Establishing a decreasing ordering of the probabilities  $p_{(1)} \geq p_{(2)} \geq \dots \geq p_{(n)}$  and associated possibility values of the distribution  $\pi_{(1)} \geq \pi_{(2)} \geq \dots \geq \pi_{(n)}$  with  $\pi_{(1)} = 1$  and  $\pi_{(n+1)} = 0$ , they built a bijective mapping defined as  $\pi_{(i)} = ip_{(i)} + \sum_{j=i+1}^n p_{(j)}$  and, inverting, obtaining  $p_{(j)} = \sum_{k=j}^n \frac{1}{k}(\pi_{(k)} - \pi_{(k+1)})$  a subject that was revisited and enhanced in subsequent publications [63-64]. Later, Yamada showed that the above mentioned transformation was one satisfying the principles of probability-possibility consistence and preference preservation [65].

Still, in the 1980's, Yager [66] had discussed specificity and entropy relative to a mathematical theory of evidence, introducing a measure of dissonance, or conflict, relative to disjoint subsets, while specificity is a measure indicating the degree to which a possibility distribution points to one (and only one) element as its manifestation – as so, a possibility distribution  $\pi$  is said to be at least as specific as another  $\pi'$  if (and only if) for each state of affairs  $s$  one has  $\pi(s) \leq \pi'(s)$ .

As Elmore *et al.* say, a possibility distribution can represent complete knowledge and complete ignorance [13]: the former, occurs when the distribution has exactly one value of 1 and all the other are 0, while complete ignorance is represented by a distribution in which all values are 1, meaning equally fully possible occurrences; the first case is also known as the dogmatic possibility distribution and the second by the vacuous distribution (e.g., [67]), the first being the extreme case of maximal specificity while the second is the opposite. Also, Higashi and Klir [16] suggested an alternative theory of information within the framework of possibility theory, and proposed a measure of possibilistic uncertainty or imprecision (U-uncertainty), conceived as a counterpart of Shannon entropy, and later proved to be the only possibilistic measure of uncertainty satisfying an axiomatic framework compatible with Shannon and Hartley's entropies [68].

Dubois and Prade discussed both contributions mentioned above and provided an interpretation of Yager's index [69], also clarifying that U-uncertainty estimates the imprecision of the focal elements. The consistency of possibility-probability transformations was studied by Delgado and Moral [70] from an axiomatic perspective, considering Zadeh formulation and Dubois and Prade criteria, highlighting a maximally specific possibility distribution, considering the usual decreasing ordering defined as  $\pi_{(i)} = \sum_{j=i}^n p_{(j)}$  for  $i = 1, \dots, n$ .

Klir has proposed a family of transformations based on the principle of uncertainty invariance – namely, the interval scale transformation – considering that uncertainty and information are interchangeable concepts [71], where non-specificity refers to ambiguity in the information and, on the other side, dissonance or conflict are associated with mutually exclusive events and Shannon entropy. Probability-possibility transformations can be conceived as mechanical manipulations,

bijjective mappings between the probabilistic and a possibilistic models. However, Sudkamp proved that there is no probability-possibility transformation, which is symmetric and expansible, and preserves second order properties concerning the relationship between two or more distributions [72], such as independence or conditionalization.

In the early 90's, Dubois, Prade and Sandri [9] made an update addressing the topic of possibility-probability transformations, stating that the principle of insufficient reason should guide the transformation of possibility into probability, while the principle of maximum specificity should operate in the opposite direction, clarifying that moving from probability to possibility in a finite setting could be expressed as finding a possibility distribution  $\pi$  dominating  $p$  satisfying the preference preservation constraint and the criterion of maximal specificity, which intends to preserve as much original information as possible. In particular, when representing measurement results one looks for maximally specific probability-possibility transformations [73].

There are several types of probability-possibility transformations and Oussalah made a detailed review on that subject providing a comparative analysis of five different transformations, regarding several criteria including specificity and preference preservation [74]. Yamada [65] clarified that if possibilities are considered a ratio scale, the bijjective transformation built by Dubois and Prade (e.g., [62-63]), can be labeled optimal, in the sense that it is the unique transformation based on equidistribution and satisfying consistency, the preference and the order preservation principles.

As Mei *et al.* summarize, there are basically three methods of probability-possibility transformation that have been suggested in the literature [75], the most common being the ratio scale method satisfying the normalization requirements of probability and possibility; the second method is the arising accumulation transformation, built by Dubois and Prade; and the third is based on Klir's method based on interval scale and information invariance.

There is a recent axiomatic characterization of the probability-possibility transformations [76], which will be used later in this work, and has already been applied in relation to an analogous situation regarding the optimal point of the weighted Gini-Simpson index [77].

### 3. Analytical Results and Inversion Procedure

In this section we present some analytical results concerning the characterization of the maximum point of weighted entropy, such as the evaluation of partial derivatives and the assessment of the dependence of the optimal point coordinates on weights, relative to monotony. Then, we will address and solve the inversion procedure.

#### 3.1. Analytical Study of the Maximum Point of Weighted Entropy

As previously mentioned, Silviu Guiaşu built what he named the principle of maximum information with no constraint [23], stating that the weighted entropy is maximum when the maximizing point is defined by Equations (2), where  $\alpha$  is the solution of the equation written below.

$$p_i^* = \exp(-(\alpha/w_i + 1)) \text{ for } i = 1, \dots, n \text{ and } \alpha: \sum_{i=1}^n \exp(-(\alpha/w_i + 1)) = 1. \quad (2)$$

Then, the maximum value is given by evaluating the quantity  $H_w^* = \alpha + \sum_{i=1}^n w_i \exp(-(\alpha/w_i + 1))$ . Next, we will make proofs of some propositions regarding the maximum point. For the partial derivatives we replace the more usual notation  $\partial f / \partial w$  for the short form  $\partial_w f$ . In the following, we are using natural logarithms, like Guiaşu did in his paper.

**Proposition 1:** For  $n \geq 3$ , and  $w_i > 0$  for  $i = 1, \dots, n$  one has  $\alpha > 0$ .

**Proof:** Consider the equivalence

$$\sum_{i=1}^n p_i^* = 1 \Leftrightarrow \sum_{i=1}^n e^{-(\alpha/w_i + 1)} = 1 \Leftrightarrow \sum_{i=1}^n e^{-(\alpha/w_i)} = e;$$

let  $w_m$  be a weight such that

$$e^{-(\alpha/w_m)} \leq e^{-(\alpha/w_k)} \text{ for } k = 1, \dots, n;$$

then, the following inequality holds:  $ne^{-(\alpha/w_m)} \leq \sum_{k=1}^n e^{-(\alpha/w_k)} = e$ .

And  $e^{-(\alpha/w_m)} \leq e/n$ ; so, for  $n \geq 3$ , and taking natural logarithms on both sides one has  $-\frac{\alpha}{w_m} \leq \ln \frac{e}{n} < 0$  what implies  $\alpha > 0$  since  $w_m > 0$  by hypothesis, and also we get that  $w_m = \min_{i=1, \dots, n} w_i$ .  $\square$

**Remark 1:** If  $n = 2$ , one has  $\alpha < 0$  as we show next. From Equations (2) we have  $e^{-(\alpha/w_1)-1} + e^{-(\alpha/w_2)-1} = 1$  or, equivalently,  $e^{-(\alpha/w_1)} + e^{-(\alpha/w_2)} = e$ ; let  $w_j$  be the weight such that  $e^{-(\alpha/w_j)} \geq e^{-(\alpha/w_k)}$ ; then, the following inequality holds  $2e^{-(\alpha/w_j)} \geq e^{-(\alpha/w_k)} + e^{-(\alpha/w_j)} = e$  and so  $e^{-(\alpha/w_j)} \geq \frac{e}{2}$  and taking logarithms we have  $-(\alpha/w_j) \geq 1 - \log 2 > 0$ , thus  $\alpha < 0$ .  $\square$

**Proposition 2:** For  $n \geq 3$ , one has  $p_i^* < e^{-1}$ .

**Proof:** From Equation (2) we have that  $p_i^* = e^{-(\alpha/w_i)-1}$ . Since  $w_i > 0$  by hypothesis and  $\alpha > 0$  as was shown in Proposition 1 we have  $e^{-(\alpha/w_i)} < 1$  and therefore we obtain  $p_i^* = e^{-(\alpha/w_i)-1} < e^{-1}$  for  $i = 1, \dots, n$ .  $\square$

**Remark 2:** Thus, the quantity  $e^{-1}$  is the supremum (least upper bound) of the maximum point coordinates.

**Proposition 3:** One has  $\partial_{w_i} \alpha = \frac{\alpha}{w_i} \left( \frac{1}{w_i} p_i^* \right) / \left( \sum_{k=1}^n \frac{1}{w_k} p_k^* \right)$ .

**Proof:** From  $\sum_{i=1}^n e^{-(\alpha/w_i)} = e$  we get

$$\partial_{w_i} \sum_{k=1}^n e^{-(\alpha/w_k)} = 0 \Leftrightarrow \sum_{k=1}^n \partial_{w_i} (e^{-(\alpha/w_k)}) = 0,$$

hence

$$\frac{\alpha}{w_i^2} e^{-(\alpha/w_i)} + \sum_{k=1}^n \left( -\frac{e^{-(\alpha/w_k)}}{w_k} \partial_{w_i} \alpha \right) = 0,$$

and evaluating  $\partial_{w_i} \alpha$  one has

$$\partial_{w_i} \alpha = \frac{\alpha}{w_i^2} e^{-(\alpha/w_i)} / \left( \sum_{k=1}^n \left( \frac{1}{w_k} e^{-(\alpha/w_k)} \right) \right).$$

Then, multiplying numerator and denominator by the quantity  $e^{-1}$  we get the result  $\partial_{w_i} \alpha = \frac{\alpha}{w_i} \left( e^{-1} \frac{1}{w_i} e^{-(\alpha/w_i)} \right) / \left( e^{-1} \left( \sum_{k=1}^n \frac{1}{w_k} e^{-(\alpha/w_k)} \right) \right) = \frac{\alpha}{w_i} \left( \frac{p_i^*}{w_i} \right) / \left( \sum_{k=1}^n \frac{p_k^*}{w_k} \right)$ .  $\square$

**Remark 3:** From the result above, we can check that  $\partial_{w_i} \alpha < \frac{\alpha}{w_i}$ .

**Proposition 4:** For  $n \geq 3$ , the partial derivatives sign is  $\partial_{w_i} p_i^* > 0$  and  $\partial_{w_j} p_i^* < 0$  if  $j \neq i$ .

**Proof:** One has

$$\partial_{w_i} p_i^* = \partial_{w_i} (e^{-(\alpha/w_i)-1}) = e^{-(\alpha/w_i)-1} \partial_{w_i} \left( -\frac{\alpha}{w_i} \right) = p_i^* \partial_{w_i} \left( -\frac{\alpha}{w_i} \right),$$

and using Proposition 3, we obtain the development  $\partial_{w_i} p_i^* = p_i^* \frac{\alpha - w_i \partial_{w_i} \alpha}{w_i^2} = p_i^* \left( \alpha - \frac{\alpha}{w_i} p_i^* / \left( \sum_{k=1}^n \frac{p_k^*}{w_k} \right) \right) / w_i^2$  from which one has  $\partial_{w_i} p_i^* = p_i^* \frac{\alpha}{w_i^2} \left( 1 - \frac{p_i^*}{w_i} / \left( \sum_{k=1}^n \frac{p_k^*}{w_k} \right) \right)$ .

Given that  $\left( \frac{p_i^*}{w_i} \right) / \left( \sum_{k=1}^n \frac{p_k^*}{w_k} \right) < 1$  and  $\alpha > 0$  (from Proposition 1), we obtain  $\partial_{w_i} p_i^* > 0$ .

Also, for  $i \neq j$  we have that

$$\partial_{w_j} p_i^* = \partial_{w_j} (e^{-(\alpha/w_i)-1}) = -e^{-(\alpha/w_i)-1} \frac{\partial_{w_j} \alpha}{w_i} \text{ and using Proposition 3 while rearranging terms we get}$$

$$\partial_{w_j} p_i^* = -\frac{\alpha}{w_j} \frac{p_j^*}{w_j} \frac{p_i^*}{w_i} / \left( \sum_{k=1}^n \frac{1}{w_k} p_k^* \right) < 0. \quad \square$$

**Remark 4:** So, we can say that for  $n \geq 3$  any optimal coordinate  $p_i^*$  increases with (the increase of) its own weight  $w_i$  and decreases with (the increase of) any other weight  $w_j$  (for  $j \neq i$ ).

**Proposition 5:** For  $n \geq 4$ , if one has  $w_j = 0$  then  $p_j^* = 0$ .

**Proof:** Since  $n \geq 4$ , from Proposition 1 we have that  $\alpha > 0$ . Then, using Equation (2) one has  $p_j^* = e^{-(\alpha/w_j+1)}$  and evaluating  $\lim_{w_j \rightarrow 0^+} p_j^*$  implies that although  $\alpha$  changes its value, yet it remains a strictly positive quantity ( $\alpha > 0$ ), because that result holds for  $n \geq 3$ ; in fact, to check this situation consider  $\sum_{i=1}^{n-1} e^{-(\alpha/w_i+1)} + e^{-(\alpha/w_n+1)} = 1$  equivalent to  $\sum_{i=1}^{n-1} e^{-\alpha/w_i} + e^{-\alpha/w_n} = e$ ; if  $\alpha \rightarrow 0$  when  $w_n \rightarrow 0$ , for the first part of the sum we would have  $\lim_{\alpha \rightarrow 0} \sum_{i=1}^{n-1} e^{-\alpha/w_i} = n - 1 > e$  if  $n \geq 4$ , thus it



can't be  $\alpha \rightarrow 0$  and we have  $a > 0$ ; then,  $\lim_{w_j \rightarrow 0^+} e^{-(\alpha/w_j+1)} = e^{-\infty} = 0$  and we can make the extension by continuity stating  $w_j = 0 \square p_j^* = 0$ , which entails that the domain becomes a face of the original simplex, what is still a simplex with lower dimension.  $\square$

**Remark 5:** The result of Proposition 5 applies for several  $w_j = 0$  as far as there remain at least three  $w_i > 0$  with  $i \neq j$  corresponding to at least three  $p_i^* > 0$ . This result is the reason why we stated in the Introduction that the case for  $w_j = 0$  for any  $j$ , or some, would be irrelevant for the scope of this article, since the event(s) matching to the outcome(s)  $\{x_j\}$  should be considered impossible and thus would be absent in the optimal point.

### 3.2. An Inversion Procedure Anchored in the Optimal Point of Weighted Entropy

The issue addressed in this section consists of the inverse problem: given a  $n$ -tuple  $(p_1^*, \dots, p_n^*)$  considered to be a solution of the maximum point of  $H_w$ , what would be a set of real positive numbers  $\{w_i\}_{i=1, \dots, n}$  that would generate that solution? We consider that the  $p_i^*$  are given in a complete setting, verifying  $0 < p_i^* < e^{-1}$  (taking into account Proposition 2) and  $\sum_{i=1}^n p_i^* = 1$ .

Ordering the optimal probabilities in a descending way like  $p_{(1)}^* \geq p_{(2)}^* \geq \dots \geq p_{(n)}^*$  and using Equations (2) with a fixed  $\alpha$  supports a parallel ordering in the values of the weights such as  $w_{(1)} \geq w_{(2)} \geq \dots \geq w_{(n)}$ , also confirmed by the study of partial derivatives in Proposition 4. In fact, an ordering procedure would not be necessary, but simplifies the reasoning and display of results, avoiding a heavier notation with permutation, and facilitates when converting to a possibility distribution.

Next, we ascribe the value  $w_{(1)} := 1$  in what could be named a fixed-pole method (e.g., [78]); one gets the simplification  $p_{(1)}^* = \exp\left(-(\alpha'/w_{(1)} + 1)\right) = e^{-1}e^{-\alpha'}$  and solving for  $\alpha'$  one has  $\alpha' = -(1 + \log p_{(1)}^*)$ ; then, we have  $p_{(2)}^* = e^{-1} \exp(-\alpha'/w_{(2)})$  equivalent to  $\log p_{(2)}^* = -1 - \frac{\alpha'}{w_{(2)}}$  and with  $\alpha' = -1 - \log p_{(1)}^*$  it follows that we get  $\log p_{(2)}^* = -1 - (-1 - \log p_{(1)}^*)/w_{(2)}$  thus obtaining  $w_{(2)} = (\log p_{(1)}^* + 1)/(\log p_{(2)}^* + 1)$ ; and, in general, we have the following relation  $w_{(j)} = (\log p_{(1)}^* + 1)/(\log p_{(j)}^* + 1)$  for  $j = 1, \dots, n$ , with  $p_{(1)}^* = \max_{i=1, \dots, n} p_i^*$ .

**Proposition 6:** The weights computed as  $w_{(i)} = (\log p_{(1)}^* + 1)/(\log p_{(i)}^* + 1)$  for  $i = 1, \dots, n$  are solution of the equation  $\sum_{i=1}^n \exp\left(-(\alpha'/w_{(i)})\right) = e$ .

**Proof:** The proof is straightforward. For a generic  $i$ , and with  $\alpha' = -1 - \log p_{(1)}^*$  one has that evaluating  $\exp\left(-(\alpha'/w_{(i)})\right)$  becomes  $\exp\left(-(-1 - \log p_{(1)}^*)/(\frac{\log p_{(1)}^* + 1}{\log p_{(i)}^* + 1})\right)$  and so  $\exp\left(-(\alpha'/w_{(i)})\right) = \exp(\log p_{(i)}^* + 1) = ep_{(i)}^*$  then  $\sum_{i=1}^n \exp\left(-(\alpha'/w_{(i)})\right) = \sum_{i=1}^n ep_{(i)}^* = e \sum_{i=1}^n p_{(i)}^* = e$ .  $\square$

**Proposition 7:** A positive linear transformation of the weights like  $u_{(i)} = cw_{(i)}$  with  $c > 0$  implies  $\alpha'' = c\alpha'$  and the optimal point coordinates remain unchanged.

**Proof:** One has  $p_{(i)}^* = \exp\left(-(\alpha'/w_{(i)} + 1)\right) = \exp\left(-(\alpha''/(cw_{(i)}) + 1)\right)$  and with  $u_{(i)} = cw_{(i)}$  we have  $\alpha'' = c\alpha'$  and  $p_{(i)}^* = \exp\left(-(\alpha''/u_{(i)} + 1)\right)$  for  $i = 1, \dots, n$ .  $\square$

### 4. Eliciting a Novel Probability-possibility Transformation

As previously mentioned in the literature review Zadeh [6] stated that the mathematical expression  $H^P(A) = -\sum_{i=1}^n \mu_A(x_i) p_i \log p_i$  evaluates the entropy of a fuzzy set  $A$  with respect to a given probability distribution  $P = (p_1, \dots, p_n)$ , where  $\mu_A(x_i)$  stands for the value of membership function concerning the occurrence with outcome  $\{x_i\}$  in relation to  $A$ . Also, the author inaugurated a theory of possibility with a basis on fuzzy sets [55], and postulated that the possibility that  $X = x_i$  denoted as  $\pi_X(x_i)$  equates the value of the membership function  $\mu_A(x_i)$ .

One sentence that highlights and expands the original formulation by Zadeh concerning possibility theory, states that a possibility distribution  $\pi_X$  is a mapping from a referential understood as a set of mutually exclusive values for the attribute  $X$  to a totally ordered scale such as the unit interval [79], and any such mapping from a set of elements, viewed as a mutually exclusive set of alternatives, can be seen as acting as an elastic restriction on the value of a single-valued variable.

We will deal in this section with Equation (3), an analogue of Equation (1), with the same domain, the simplex  $\Delta^{n-1} = \{(p_1, \dots, p_n) : p_i \geq 0, \sum_{i=1}^n p_i = 1\}$ , and the possibility distribution values considered real numbers in the interval  $0 < \pi_i \leq 1$ , associated with the normalization condition  $\max_{i=1, \dots, n} \pi_i = 1$  for some  $i$ .

$$H_\pi = -\sum_{i=1}^n \pi_i p_i \log p_i. \quad (3)$$

Therefore, observing Equation (3), we see that the syntax of Equation (1) holds, but the semantics has changed somewhat.

In fact, if one refers to the quantity  $g_i = -\log p_i = \log(1/p_i) > 0$  when  $p_i < 1$ , as the information gain associated with the occurrence of the matching event (e.g., [52]), then Shannon entropy evaluates the uncertainty associated with a discrete and finite random variable  $X$ ; now, in Equation (3) one has another quantity,  $f_i = -\pi_i \log p_i$ , a fuzzy information gain verifying  $f_i = \pi_i g_i \leq g_i$  as  $0 < \pi_i \leq 1$ , modulated by another level of imprecision concerning incomplete knowledge about the situation, for instance a lack of accuracy. For example, when throwing a dice with the dots faded, making it difficult or impossible to recognize the outcome with naked eye (but, presumably, not with a lens or specific test), that can be considered a kind of incomplete knowledge coupled with ambiguity: rolling a dice means a certain outcome, yet if one cannot identify it properly then becomes a matter of incomplete knowledge or fuzziness.

Concomitantly, we will not refer to  $H_\pi$  as a measure of uncertainty, as we consider that the information gain keeps exclusively associated with randomness, and reducing inaccuracy amplifies the gain; so, we can refer to Equation (3) as ‘possible information’ using an analogy with the ‘useful information’ mentioned in the literature review concerning weighted entropy. And obviously, one has  $H_\pi \leq H$  for the same probability distribution and equality is only attained in the case  $\pi_i = 1$  for all singletons with strictly positive probability. However, Equation (3) should not be confused and misinterpreted with fuzzy entropy in other senses, like the concept of a non-probabilistic entropy of fuzzy sets (e.g., [80]).

We had adopted Equation (3) as a synthesis of a possibility-weighted entropy, where we denote for short  $\pi_i := \pi_X(x_i)$  associated with a complete ordering such as  $\pi_{(1)} \geq \dots \geq \pi_{(n)} \geq \pi_{(n+1)} = 0$ , with  $\pi_{(1)} = 1$  meaning that the correspondent outcome  $\{x_1\}$  of an elementary event is considered fully possible which is the normalization condition, associated with the parallel probability ordering  $p_{(1)}^* \geq p_{(2)}^* \geq \dots \geq p_{(n)}^*$ .

Also, one must keep in mind the consistence principle previously mentioned by Dubois and Prade (e.g., [54], [79]), stating that one must have  $\pi_{(i)} \geq p_{(i)}^*$  for  $i = 1, \dots, n$ , and remind that Equation (3) can be considered a framework for dealing with a joint description of a random variable and a fuzzy variable, associating the two uncertainty systems of probability and possibility, as other methods do (e.g., [81]).

Let us remember the results presented before in the propositions related to the weights, and replace the notation with the most usual in probability-possibility transformations changing  $w_{(i)}$  by  $\pi_{(i)}$  for  $i = 1, \dots, n$ , with  $\pi_{(1)} = 1$  for normalization and adapting Proposition 6, we get the values  $\pi_{(j)}$  defined by the following equation(s):

$$\pi_{(i)} = (\log p_{(1)}^* + 1) / (\log p_{(i)}^* + 1) \text{ for } i = 1, \dots, n. \quad (4)$$

## 5. Discussion of Results

In this section, we will provide two types of results: theoretical, relative to checking axioms for an admissible probability-possibility transformation; and empirical, scrutinizing numeric examples. Last, we will make general comments and suggestions on this subject, including possible further research.

### 5.1. Checking Axioms

Jin *et al.* presented a set of four axioms (see [76]), regarding an admissible probability-possibility transformation, that we will check, relative to the problem at stake associated with Equations (4).

- A1) Bijectivity. If  $T: p \square \pi$  is such a transformation,  $T$  should be injective so that there is an inverse transformation  $T^{-1}: \pi \square p$ ; the domain of  $T$  is the open interval  $]0, e^{-1}[$  (from Proposition 2), with a defined  $p_{(1)}^* = \max_{i=1, \dots, n} p_i^*$  from which one sets  $\pi_{(1)} = 1$  for normalization.

From Equations 4, we have  $\pi_{(j)} = (\log p_{(1)}^* + 1) / (\log p_{(j)}^* + 1)$  and manipulating the expression one gets  $p_{(j)}^* = f(\pi_{(j)}) = e^{-1} \exp((\log p_{(1)}^* + 1) / \pi_{(j)})$ ; thus, for a given  $p_{(1)}^*$ , it follows that  $p_{(j)}^* = f(\pi_{(j)})$  is defined as a one-to-one correspondence with  $0 < \pi_{(j)} \leq 1$  and  $0 < p_{(j)}^* < e^{-1}$ ; yet,  $T^{-1}$  is not surjective concerning the whole unit interval  $[0, 1]$ , but stands only in the restriction mentioned, the open interval  $]0, e^{-1}[$ , and bijectivity only applies there;

- A2) Concerning co-monotonicity, one should observe that the following equivalence holds:

$$p_{(i)}^* \geq p_{(j)}^* \square \pi_{(i)} \geq \pi_{(j)}; \text{ one has } \pi_{(i)} \geq \pi_{(j)} \square \frac{\log p_{(1)}^* + 1}{\log p_{(i)}^* + 1} \geq \frac{\log p_{(1)}^* + 1}{\log p_{(j)}^* + 1} \text{ and inverting both members, we}$$

$$\text{obtain } \frac{\log p_{(i)}^* + 1}{\log p_{(1)}^* + 1} \leq \frac{\log p_{(j)}^* + 1}{\log p_{(1)}^* + 1} \text{ from what one can multiply by } \log p_{(1)}^* + 1 < 0 \text{ (because } p_{(1)}^* < e^{-1}),$$

$$\text{then obtaining } \log p_{(i)}^* + 1 \geq \log p_{(j)}^* + 1 \text{ and } p_{(i)}^* \geq p_{(j)}^*;$$

- A3) Consistency, entailing that for  $j = 1, \dots, n$  one should have  $\pi_{(j)} \geq p_{(j)}^*$ ; then, it should hold

$$\text{that } \frac{\log p_{(1)}^* + 1}{\log p_{(j)}^* + 1} \geq p_{(j)}^* \square p_{(j)}^* (\log p_{(j)}^* + 1) \geq \log p_{(1)}^* + 1 \text{ with } p_{(1)}^*, p_{(j)}^* \in ]0, e^{-1}[; \text{ one can research}$$

the behavior of  $f(p_{(j)}^*) = p_{(j)}^* (\log p_{(j)}^* + 1)$  in the interval  $0 < p_{(j)}^* < e^{-1}$ ;  $f(p_{(j)}^*)$  has a single minimum point at  $p_{(j)}^* = e^{-2}$  with value  $f(e^{-2}) = -e^{-2}$ ; so, if we want to ensure that the inequality  $p_{(j)}^* (\log p_{(j)}^* + 1) \geq \log p_{(1)}^* + 1$  stands, a sufficient condition is that we have  $-e^{-2} \geq \log p_{(1)}^* + 1 \square \log p_{(1)}^* \leq -1 - e^{-2}$  and then  $p_{(1)}^* \leq \exp(-1 - e^{-2})$ ; so, there is a threshold relative to  $p_{(1)}^*$  for ensuring consistency of the transformation; previously, we had already the condition  $p_{(1)}^* < e^{-1} \cong 0.36788$  but now that quantity is lessened about 0.047 to ensure axiom A3. We denote that threshold as  $\tau = \exp(-1 - e^{-2}) \cong 0.32131$ ;

- A4) Support preserving, meaning  $p_{(i)}^* = 0$  if and only if  $\pi_{(i)} = 0$ . From proposition 5 we know that, for  $n \geq 4$ , the following implication holds  $\pi_{(i)} = 0 \square p_{(i)}^* = 0$ ; then, with  $\pi_i = (\log p_{(1)}^* + 1) / (\log p_{(i)}^* + 1)$  we can evaluate the extension by continuity with the limit

$$\lim_{p_{(i)}^* \rightarrow 0^+} \frac{\log p_{(1)}^* + 1}{\log p_{(i)}^* + 1} = 0 \text{ and infer } p_{(i)}^* = 0 \square \pi_{(i)} = 0.$$

Summarizing, the axioms were verified with some restrictions: concerning A1, when we move from probability to possibility the domain is the open interval  $]0, e^{-1}[$  and therefore the inverse transformation, although injective, is not surjective in relation to the whole unit interval  $[0, 1]$ ; referring to A3, for assuring consistency one has to observe a sufficient condition relative to a threshold more restrictive than the least upper bound for the optimal probabilities, meaning  $\max_{i=1, \dots, n} p_i^* < \exp(-1 - e^{-2}) = \tau$ .

## 5.2. Numerical Examples

Below, we present detailed numerical examples that help illustrate and discuss the probability-possibility transformation outlined in this paper; in any case, the examples have a small dimension  $n = 7$ , suitable to convey a qualitative appreciation and analysis. Two other examples aim to complement and highlight the relevance of the threshold mentioned when discussing axiom A3.

In the first case, we computed the optimal probability point of weighted entropy given the weights, using Equations (2) outlined by Guiaşu. In the second example, the probability distribution was imported from a numerical example previously discussed, concerning the weighted Gini-

Simpson index [77]. Then, using the probability distribution(s) as input(s), we computed the novel possibility distribution with the procedure outlined in this article in Equations (4).

Regarding the first two examples (and also using the correspondent probability distribution as input) other four different possibility distributions mentioned in the literature were evaluated to be compared with the new one, namely:

- The one derived from the weighted Gini-Simpson index (see [77]) denoted  $\pi_{(i)}^0$  and computed like  $\pi_{(i)}^0 = (1 - 2p_{(1)}^*) / (1 - 2p_{(i)}^*)$ ;
- The so-called 'optimal transformation', mentioned in the literature review and considered well-accepted (e.g., [13], [79]) defined as  $\pi'_{(i)} = \sum_{j=i}^n p_{(j)}^*$ , which is considered satisfying criteria of consistency, ordinal faithfulness and maximal specificity;
- The 'ratio scale transformation', quite well known and mentioned (e.g., [74], [79], [81]), denoted  $\pi''_{(i)} = p_{(i)}^* / p_{(1)}^*$  with  $p_{(1)}^* = \max_{i=1, \dots, n} p_i^*$ ;
- And, last, the 'arising accumulation transformation' (e.g., [76]), derived by Dubois and Prade [62-63], calculated with the expression  $\pi'''_{(i)} = ip_{(i)}^* + \sum_{j=i+1}^n p_{(j)}^*$ .

Still, we will compare all the possibility distributions evaluated with three criteria or measures well known from the literature:

1. The heuristic weak probability-possibility consistency principle of Zadeh (e.g., [55], [62]), computed like  $\gamma = \mathbf{p}^* | \boldsymbol{\pi}$ , meaning the inner product of the vectors of probabilities and of the possibility distribution(s) at stake, which was discussed with consistency axioms and proved to be a solution for predicting  $\boldsymbol{\pi}$  given  $\mathbf{p}^*$  [70]; one has that  $0 < \gamma \leq 1$ , and the smaller the value (close to 0) the more orthogonal are the vectors, while near the value 1 approximates the 'vacuous distribution' ( $\pi_{(i)} = 1$  for  $i = 1, \dots, n$ ), reflecting trivial possibility.
2. A measure of specificity (e.g., [13], [82]), evaluated as  $\sigma = 1 - \sum_{j \neq 1} \pi_{(j)} / (n - 1)$  coupled with the condition  $\pi_{(1)} = 1$ , assessing how a possibility distribution is countering uncertainty, the maximal specificity ( $\sigma = 1$ ) being attained when a singleton is totally possible and all the other impossible, the so-called 'dogmatic distribution' ( $\pi_{(j)} = 1$  and  $\pi_{(i)} = 0$  if  $i \neq j$ ); on the opposite, the minimal specificity is achieved with the 'vacuous distribution', all singletons being totally possible, implying maximum uncertainty and ignorance and thus null specificity ( $\sigma = 0$ );
3. Last, it will be evaluated the total possibilistic uncertainty (e.g., [83-84]) here denoted  $\vartheta = \sum_{i=2}^n (\pi_{(i)} - \pi_{(i+1)}) \log_2(i^2 / \sum_{j=1}^i \pi_{(j)})$  with  $\pi_{(n+1)} = 0$ , which can be considered inversely related with specificity; once more, the maximum value being  $\vartheta = \log_2 n$  (if  $n = 7$ ,  $\log_2 7 \cong 2.8074$ ) attained with the 'vacuous distribution', and the minimum value (equal to 0) relative to the 'dogmatic distribution'.

### 5.2.1. Numerical example computed with the optimal point of weighted entropy

In this first example, seen in Table 1, given ordinal weights in decreasing order  $w_{(1)} = 7, w_{(2)} = 6, \dots, w_{(7)} = 1$ , the maximum point of weighted entropy was evaluated with Equations (2) and we can see that  $\mathbf{p}^*$  seems decreasingly balanced, ranging from  $p_{(1)}^* = 0.23$  to  $p_{(7)}^* = 0.02$ ; the calculated probabilities vary by about 11:1, while the original weights vary by a ratio of 7:1. The possibility distribution computed with Equations (4) is denoted  $\pi_{(i)}$  and the others mentioned above ( $\pi_{(i)}^0, \pi'_{(i)}, \pi''_{(i)}, \pi'''_{(i)}$ ) were computed with the same probability distribution as input, for comparison purposes.





$\gamma$	0.45	0.69	0.60	0.65	0.80
$\sigma$	0.86	0.52	0.72	0.65	0.44
$\vartheta$	0.85	1.96	1.55	1.75	2.26

<sup>1</sup>Numbers are presented rounded up with two decimal places for readability, although they were computed to fifteen digits.

From Table 2, one can see that  $\pi_{(i)}$  performs differently from the previous example, when compared with all the other four possibility distributions.

First, we should note that  $p_{(1)}^* = 0.32$  is very near the threshold that was identified concerning axiom A3 ( $\tau \cong 0.32131$ ); yet, the figures still support general consistency as it should. When we move from  $p_{(1)}^* = 0.32$  to  $p_{(2)}^* = 0.23$  there is a decline of about 28%, but, correspondingly, from  $\pi_{(1)} = 1.00$  to  $\pi_{(2)} = 0.30$  there is a sharp decline of 70%; thus  $\pi_{(i)}$  shows the more orthogonal with the vector of probabilities when comparing with all the other possibility distributions using the criterium of Zadeh with the lowest value ( $\gamma = 0.45$ ); also, it has the highest specificity index ( $\sigma = 0.86$ ), what derives from the same abrupt transition between  $\pi_{(1)}$  and all the other possibility values of the distribution; last, and correspondingly, it has by far the least uncertainty measure  $\vartheta = 0.85$ .

Then, it seems plausible consider that when one has  $p_{(1)}^* \cong \tau$ , our new probability-possibility transformation generating  $\pi_{(i)}$  reveals a kind of abrupt behavior, regarding the possibility values following next to the maximum.

Rechecking that situation with a similar example built with that goal: in Table 3 we can see a similar probability distribution, where the maximum value  $p_{(1)}^* = 0.32$  is the same as the previous example, but next follows  $p_{(2)}^* = 0.30$ , meaning a decrease of less than 10% comparing with  $p_{(1)}^*$ , but the correspondent possibility values go from  $\pi_{(1)} = 1$  to  $\pi_{(2)} = 0.68$  with a decline of 32%.

So, it seems that when  $p_{(1)}^*$  is near the threshold  $\tau \cong 0.32131$  there will be an abrupt transition relative to the next term of the possibility distribution, even if the probability values are quite near; yet,  $\pi_{(2)}$  shows the same transition in this example, and the other possibility distributions reckoned in the case seem considerably more conservative, namely  $\pi_{(i)}^0$  and  $\pi_{(i)}'''$ .

**Table 3.** The input probability distribution is built to recheck the anomaly near the threshold. The possibility distributions are reckoned as usual

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$
$p_{(i)}^*$	0.32	0.30	0.18	0.10	0.05	0.03	0.02
$\pi_{(i)}$	1.00	0.68	0.20	0.11	0.07	0.06	0.05
$\pi_{(i)}^0$	1.00	0.90	0.56	0.45	0.40	0.38	0.38
$\pi_{(i)}'$	1.00	0.68	0.38	0.20	0.10	0.05	0.02
$\pi_{(i)}''$	1.00	0.94	0.56	0.31	0.16	0.09	0.06
$\pi_{(i)}'''$	1.00	0.98	0.74	0.50	0.30	0.20	0.14

<sup>1</sup>Numbers are presented rounded up with two decimal places for readability, although they were computed to fifteen digits.

5.2.3. Numerical example used to check consistency when the threshold  $\tau$  is violated

Here we use a numerical example where  $p_{(1)}^* = 0.34$  meaning it is below the absolute constraint  $e^{-1}$  but above the threshold previously identified when discussing axiom A3, namely  $\tau \cong 0.32131$ . Once more, given the probability distribution as input, the values of the new possibility distribution were computed using Equations (4). The results can be seen in Table 4.

**Table 4.** The input probability distribution is built to check the threshold relative to axiom A3. The possibility distributions are reckoned as usual

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$
$p_{(i)}^*$	0.34	0.29	0.18	0.10	0.04	0.03	0.02

$\pi_{(i)}$	1.00	0.33	0.11	0.06	0.04	0.03	0.03
$\pi_{(i)}^0$	1.00	0.76	0.50	0.40	0.35	0.34	0.33
$\pi'_{(i)}$	1.00	0.66	0.37	0.19	0.09	0.05	0.02
$\pi''_{(i)}$	1.00	0.85	0.53	0.29	0.12	0.09	0.06
$\pi'''_{(i)}$	1.00	0.95	0.73	0.49	0.25	0.20	0.14

<sup>1</sup>Numbers are presented rounded up with two decimal places for readability, although they were computed to fifteen digits.

From Table 4, one can see that in fact  $\pi_{(3)} < p_{(3)}^*$  and  $\pi_{(4)} < p_{(4)}^*$  and so the threshold shows to be relevant for assessing axiom A3; in this case, the possibility distribution does not verify the consistency principle which would demand  $\pi_{(i)} \geq p_{(i)}^*, \forall i$ . The noticeable flaws occur in the middle of the distribution, and none of the other possibility distributions show this anomaly.

### 5.3. General Comments and Further Research

As Mei stated, probability comes with randomness while possibility comes with fuzziness [86], and understanding randomness, fuzziness and unknown-uncertainty, and their related theories, keeps evolving [81]. Also, probability-possibility transformations is a field of research that goes on with new developments (e.g. [87-88]), as well as appear new frameworks dealing with weighted entropic models [89] or random and epistemic uncertainties (e.g. [90]), including portfolio selection based on fuzzy weighted entropy [91].

The inversion procedure outlined in this paper, anchored in the optimal point of weighted entropy and using the normalization condition  $\max_{i=1,\dots,n} w_i = 1$  is feasible and appeals naturally for considering a possibility distribution associated with the probability distribution at stake, which was established using Equations (4) and further discussed.

From the theoretical results, it follows that the main limitation of this method is the one linked to the intrinsic restriction on the evaluation of optimal coordinates, bounded above by the *supremum* (least upper bound) value  $\exp(-1)$  if  $n \geq 3$ . Still, when using the consistency criterium defined by Dubois and Prade implying  $\pi_{(i)} \geq p_{(i)}^* \forall i$ , one has another constraint concerning a threshold, entailing  $\max_{i=1,\dots,n} p_i^* \leq \tau = \exp(-1 - e^{-2})$  for consistency. However, it should be reminded that a restriction of the type was already mentioned concerning the *ratio scale* transformation (e.g., [79]), and it could be also noted that Lofti Zadeh in his last papers emphasized that restriction is the principle which sustains information ([92-93]).

It is easily seen that one could retrieve standardized weights from the results of the procedure making  $w'_i = \pi_i / \sum_{j=1}^n \pi_j$  obtaining adimensional numbers verifying intrinsically  $0 < w'_i < 1$  and  $\sum_{i=1}^n w'_i = 1$ , or else any other form of linear positive scaling as  $w''_i = c\pi_i$  with  $c > 0$ , as by Proposition 7 the optimal point coordinates are insensitive to such a transformation. And from there we can move forward addressing an utility structure anchored in values for goals, a subject of research that goes on, including model selection in artificial intelligence using fuzzy weighted entropy [94] or dynamic discrete choice where weighted entropy is used to assess information cost [95].

Concerning the main limitation mentioned, relative to the barrier defined by the *supremum* of the maximum point coordinates clarified in Proposition 2, it cannot be overcome within the framework of weighted Shannon entropy. But we can move into another related one – mentioned as an expected utility and weighted entropy model – also defined with domain in the usual simplex, herein denoted as  $E_w = H_w + \sum_{i=1}^n w_i p_i$  where it is shown that the optimal coordinates verify  $0 < p_i^* < 1$  ([43], [96-97]), which was previously used for assessing and discussing landscape mosaic composition with different habitats, within the field of landscape ecology.

**Author Contributions:** JPC: Conceptualization, Methodology, Writing – original draft. VM: Formal analysis; Writing – review & editing. Both authors have read and agreed to the current version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data is contained within the article.

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

## References

1. Belis, M.; Guiaşu, S. A quantitative-qualitative measure of information in cybernetic systems. *IEEE T. Inf. Theory* **1968**, *14*(4), 593-594. <http://doi.org/10.1109/TIT.1968.1054185>
2. Von Neumann, J.; Morgenstern, O. *Theory of Games and Economic Behavior* 3rd Ed.; Princeton University Press: Princeton, NJ, USA, 1953; pp. 15-30.
3. Debreu, G. *Theory of Value –An Axiomatic Analysis of Economic Equilibrium*. Yale University Press: NewHaven and London, 1959; pp. 50-73.
4. Longo, G. The communication problem: the statistical, the semantic and the effectiveness viewpoints — sources having utilities and the useful entropy. In *Quantitative-Qualitative Measure of Information*; International Centre for Mechanical Sciences; Springer: Vienna, Austria, 1972; Volume 138, pp. 17-26. [http://doi.org/10.1007/978-3-7091-2796-4\\_2](http://doi.org/10.1007/978-3-7091-2796-4_2)
5. Monroe, T.; Beruvides, M.; Tercero-Gómez, V. Derivation and application of the subjective-objective probability relationship from entropy: the entropy decision risk model (EDRM). *Systems* **2020**, *8*, 46. <http://doi.org/10.3390/systems8040046>
6. Zadeh, L. A. (1968). Probability measures of fuzzy events. *J. Math. Anal. Appl.* **1968**, *23*(2), 421-427. [http://doi.org/10.1016/0022-247X\(68\)90078-4](http://doi.org/10.1016/0022-247X(68)90078-4)
7. Zadeh, L. A. (1965). Fuzzy sets. *Inform. Control* **1965**, *8*, 338-353. [http://doi.org/10.1016/S0019-9958\(65\)90241-X](http://doi.org/10.1016/S0019-9958(65)90241-X)
8. Dubois, D.; Prade, H. Reasoning and learning in the setting of possibility theory - overview and perspectives. *Int. J. Approx. Reason.* **2024**, *171*, 109028. <http://doi.org/10.1016/j.ijar.2023.109028>
9. Dubois, D.; Prade, H.; Sandri, S. On possibility/probability transformations. In *Fuzzy Logic*; Lowen, R., Roubens, M., Eds; Theory and Decision Library; Springer: Dordrecht, Netherlands, 1993; Volume 12, pp. 103-112. [https://doi.org/10.1007/978-94-011-2014-2\\_10](https://doi.org/10.1007/978-94-011-2014-2_10)
10. Denoeux, T. Reasoning with fuzzy and uncertain evidence using epistemic random fuzzy sets: general framework and practical models. *Fuzzy Sets Syst.* **2023**, *453*, 1-36. <https://doi.org/10.1016/j.fss.2022.06.004>
11. Dubois, D., Prade, H. Possibility theory. In *Granular, Fuzzy, and Soft Computing. Encyclopedia of Complexity and Systems Science Series*; Lin, T.Y., Liau, C.J., Kacprzyk, J. Eds; Springer: New York, USA, 2023, pp. 859-876. [https://doi.org/10.1007/978-1-0716-2628-3\\_413](https://doi.org/10.1007/978-1-0716-2628-3_413)
12. Dubois, D.; Pap, E.; Prade, H. Hybrid probabilistic-possibilistic mixtures and utility functions. In *Preferences and Decisions under Incomplete Knowledge*; Fodor, J., De Baets, B., Perny, P., Eds; Studies in Fuzziness and Soft Computing; Physica: Heidelberg, Germany, 2000; Volume 51, pp. 51-73. [https://doi.org/10.1007/978-3-7908-1848-2\\_4](https://doi.org/10.1007/978-3-7908-1848-2_4)
13. Elmore, P.; Anderson, D.; Petry, F. Evaluation of heterogeneous uncertain information fusion. *J. Amb. Intel. Hum. Comp.* **2020**, *11*(6), 799-811. <http://doi.org/10.1007/s12652-019-01320-3>
14. Le Carrer, N.; Ferson, S. Beyond probabilities: a possibilistic framework to interpret ensemble predictions and fuse imperfect sources of information. *Q. J. Roy. Meteor. Soc.* **2021**, *147*, 3410-3433. <http://doi.org/10.1002/qj.4135>
15. Hartley, R. V. L. Transmission of information. *Bell Syst. Tech. J.* **1928**, *7*(3), 535-563. <http://doi.org/10.1002/j.1538-7305.1928.tb01236.x>
16. Higashi, M.; Klir, G. J. Measures of uncertainty and information based on possibility distributions. *Int. J. Gen. Syst.* **1982**, *9*, 43-58. <https://doi.org/10.1080/03081078208960799>
17. Shannon, C. E.. A mathematical theory of cryptography. Alcatel-Lucent 1945 <https://www.iacr.org/museum/shannon/shannon45.pdf> (assessed on 01 August 2024).
18. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379-423. <http://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
19. Kapur, J. N. A comparative assessment of various measures of entropy. *J. Inf. Optim. Sci.* **1983**, *4*(3), 207-232. <http://dx.doi.org/10.1080/02522667.1983.10698762>
20. Csiszár, I. Axiomatic characterizations of information measures. *Entropy* **2008**, *10*(3), 261-273. <http://doi.org/10.3390/e10030261>
21. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*(4), 620-630. <http://doi.org/10.1103/PhysRev.106.620>
22. Pal, N. R.; Pal, S. K. Entropy: a new definition and its applications. *IEEE T. Syst. Man Cyb.* **1991**, *21*(5), 1260-1270. <http://doi.org/10.1109/21.120079>
23. Guiaşu, S. Weighted entropy. *Rep. Math. Phys.* **1971**, *2*(3), 165-179. [http://doi.org/10.1016/0034-4877\(71\)90002-4](http://doi.org/10.1016/0034-4877(71)90002-4)

24. Skala, H. J. Remarks on semantic information. In *Information, Inference and Decision*; Menges, G., Ed.; Theory and Decision Library; Springer: Dordrecht, Netherlands, 1974; Volume 1, pp. 181-188. [http://doi.org/10.1007/978-94-010-2159-3\\_10](http://doi.org/10.1007/978-94-010-2159-3_10)
25. Bouchon, B. Useful information and questionnaires, *Inform. Control* **1976**, 32, 368-378. [http://doi.org/10.1016/S0019-9958\(76\)90279-5](http://doi.org/10.1016/S0019-9958(76)90279-5)
26. Sharma, B. D.; Mitter, J.; Mohan, M. On measures of “useful” information. *Inform. Control* **1978**, 39(3), 323-336. [http://doi.org/10.1016/S0019-9958\(78\)90671-X](http://doi.org/10.1016/S0019-9958(78)90671-X)
27. Aggarwal, N. L.; Picard, C. F. Functional equations and information measures with preference. *Kybernetika* **1978**, 14(3), 174-181. <http://www.kybernetika.cz/content/1978/3/174> (accessed on 20 August 2024)
28. Dial, G.; Taneja, I. J. On weighted entropy of type  $(\alpha, \beta)$  and its generalizations. *Aplikace Matematiky* **1981**, 26(6), 418-425. : <http://dml.cz/dmlcz/103931> (accessed on 10 August 2024)
29. Batty, M. Cost, Accessibility and weighted entropy. *Geogr. Anal.* **1983**, 15(3), 256-267. <http://doi.org/10.1111/j.1538-4632.1983.tb00786.x>
30. Nawrocki, D.N.; Harding, W.H. State-value weighted entropy as a measure of investment risk. *Appl. Econ.* **1986**, 18(4), 411-419. <http://doi.org/10.1080/00036848600000038>
31. Guiaşu, S. Grouping data by using the weighted entropy. *J. Stat. Plan. Infer.* **1986**, 15, 63-69. [http://doi.org/10.1016/0378-3758\(86\)90085-6](http://doi.org/10.1016/0378-3758(86)90085-6)
32. Jumarie, G. A concept of observed weighted entropy and its application to pattern recognition. *Pattern Recogn. Lett.* **1987**, 5, 191-194. [http://doi.org/10.1016/0167-8655\(87\)90062-6](http://doi.org/10.1016/0167-8655(87)90062-6)
33. Criado, F.; Gachechiladze, T. Entropy of fuzzy sets. *Fuzzy Sets Syst.* **1997**, 88, 99-107. [http://doi.org/10.1016/S0165-0114\(96\)00073-5](http://doi.org/10.1016/S0165-0114(96)00073-5)
34. Cios, K.J.; Pedrycz, W.; Swiniarski, R.W. Fuzzy sets. In *Data Mining Methods for Knowledge Discovery*; The Springer International Series in Engineering and Computer Science; Springer: Boston, MA., USA, 1998; Volume 458, pp. 73-129. [http://doi.org/10.1007/978-1-4615-5589-6\\_3](http://doi.org/10.1007/978-1-4615-5589-6_3)
35. Simpson, E. Measurement of diversity. *Nature* **1949**, 163, 688. <http://doi.org/10.1038/163688a0>
36. Casquilho, J. P.; Mena-Matos, H. On the optimal point of the weighted Simpson index. *Mathematics* **2024**, 12, 507. <http://doi.org/10.3390/math12040507>
37. Hu, W.; Starr, A. G.; Zhou, Z.; Leung, A. Y. T. A systematic approach to integrated fault diagnosis of flexible manufacturing systems. *Int. J. Mach. Tool. Manu.* **2000**, 40(11), 1587-1602. [http://doi.org/10.1016/S0890-6955\(00\)00016-X](http://doi.org/10.1016/S0890-6955(00)00016-X)
38. Guiaşu, R.C.; Guiaşu, S. Conditional and weighted measures of ecological diversity. *Int. J. Uncertain. Fuzz.* **2003**, 11(3), 283-300. <http://doi.org/10.1142/S0218488503002089>
39. Srivastava, A.; Maheshwari, S. A new weighted information generating function for discrete probability distributions. *Cybern. Inf. Technol.* **2011**, 11(4), 24-30. [http://cit.iict.bas.bg/CIT\\_2011/v11-4/Srivastava-Maheshwari-24-30.pdf](http://cit.iict.bas.bg/CIT_2011/v11-4/Srivastava-Maheshwari-24-30.pdf) (accessed on 03 August 2024)
40. Hooda, D. S.; Sharma, D. K. Generalized ‘useful’ information generating functions. *J. Appl. Math. Inform.* **2009**, 27(3-4), 591-601. <http://koreascience.or.kr/article/JAKO200919463950671.pdf> (accessed on 05 August 2024)
41. Ebanks, B. R. Weighted entropies. *Centr. Eur. J. Math.* **2010**, 8(3), 602-615 <http://doi.org/10.2478/s11533-010-0026-3>
42. Casquilho, J.A.P. Ecomosaico florestal: composição, índices de informação e abdução. *Rev. Árvore* **2012**, 36(2), 321-329. <http://doi.org/10.1590/S0100-67622012000200013>
43. Casquilho, J.; Neves, M.; Rego, F. C. Extensões da função de Shannon e equilíbrios de proporções – uma aplicação ao mosaico de paisagem. *An. Inst. Sup. Agron.* **1997**, 46, 77-99. <http://hdl.handle.net/10400.5/16181> (accessed on 30 August 2024)
44. Patsakis, C.; Mermigas, D.; Pirounias, S.; Chondrokoukis, G. The role of weighted entropy in security quantification. *Int J. Inf. Electron. Eng.* **2013**, 3(2), 156-159. <http://doi.org/10.7763/IJIEE.2013.V3.288>
45. Singer, G.; Anuar, R.; Ben-Gal, I. A weighted information-gain measure for ordinal classification trees. *Expert Syst. Appl.* **2020**, 152, 113375. <http://doi.org/10.1016/j.eswa.2020.113375>
46. Gkelsinis, T.; Karagrigoriou, A. Theoretical aspects on measures of directed information with simulations. *Mathematics* **2020**, 8(4), 587. <http://doi.org/10.3390/math8040587>
47. Zhou, J.; Chen, K.; Liu, J. A clustering algorithm based on the weighted entropy of conditional attributes for mixed data. *Concurr. Comp.- Pract. E.* **2021**, 33(17), e6293. <http://doi.org/10.1002/cpe.6293>
48. Li, Y-M.; Xiang, Y. Detecting disease association with rare variants using weighted entropy. *J. Genet.* **2023**, 102, 34. <http://doi.org/10.1007/s12041-023-01432-y>
49. Li, Y.; Zhang, S.; Xu, T.; Zhu, M.; Zhu, Q. Relatively important node identification for cyber-physical power systems based on relatively weighted entropy. *Int. J. Elec. Power* **2024**, 160, 110050. <http://doi.org/10.1016/j.ijepes.2024.110050>
50. Bouchon-Meunier, B. ; Marsala, C. Entropy and monotonicity in artificial intelligence. *Int. J. Approx. Reason.* **2020**, 124, 111-122. <http://doi.org/10.1016/j.ijar.2020.04.008>



51. Aggarwal, M. An entropy framework for randomness and fuzziness. *Expert Syst. Appl.* **2024**, *243*, 122431 . <http://doi.org/10.1016/j.eswa.2023.122431>
52. Aggarwal, M. Bridging the gap between probabilistic and fuzzy entropy. *IEEE T. Fuzzy Syst.* **2020**, *28*(9), 2175-2184. <http://doi.org/10.1109/TFUZZ.2019.2931232>
53. Cox, R. T. Probability, frequency and reasonable expectation. *Am. J. Phys.* **1946**, *14*(1), 1-13. <https://doi.org/10.1119/1.1990764>
54. Dubois, D., Prade, H. Possibility theory and its applications: Where do we stand? In *Springer Handbook of Computational Intelligence*; Kacprzyk, J., Pedrycz, W., Eds; Springer: Berlin, Heidelberg, Germany, 2015; pp. 31-60. [https://doi.org/10.1007/978-3-662-43505-2\\_3](https://doi.org/10.1007/978-3-662-43505-2_3)
55. Zadeh, L. A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* **1978**, *1*, 3-28. [http://doi.org/10.1016/S0165-0114\(99\)80004-9](http://doi.org/10.1016/S0165-0114(99)80004-9)
56. Gaines B. R.; Kohout L Possible automata. In Proceedings of the International Symposium on Multiple-Valued Logic, Bloomington, Indiana, USA, 13–16 May 1975. IEEE Press, pp. 183–196. <https://cspages.ucalgary.ca/~gaines/reports/SYS/PA75/PA75.pdf> (accessed on 15 August 2024)
57. Zio, E.; Pedroni, N. Literature review of methods for representing uncertainty. *Cahiers de Sécurité Industrielle* **2013**, *3*, 1-49. <http://doi.org/10.57071/124ure>
58. Dubois, D. Possibility theory and statistical reasoning. *Compu. Stat. Data An.* **2006**, *51*(1), 47-69. <https://doi.org/10.1016/j.csda.2006.04.015>
59. Denœux, T.; Dubois, D.; Prade, H. Representations of uncertainty in artificial intelligence: probability and possibility. In *A Guided Tour of Artificial Intelligence Research*; Marquis, P., Papini, O., Prade, H., Eds; Springer: Cham, Switzerland, 2020; pp. 69-117. [https://doi.org/10.1007/978-3-030-06164-7\\_3](https://doi.org/10.1007/978-3-030-06164-7_3)
60. Dubois, H.; Prade, D. *Fuzzy Sets and Systems: Theory and Applications*; Academic Press, Inc. Chestnut Hill, MA, USA; Academic Press Ltd: London, UK. 1980.
61. Dubois, H.; Prade, D. Evidence measures based on fuzzy information. *Automatica*, **1985**, *21*(5), 547-562. [http://doi.org/10.1016/0005-1098\(85\)90004-4](http://doi.org/10.1016/0005-1098(85)90004-4)
62. Dubois, D.; Prade, H. On several representations of an uncertain body of evidence. In *Fuzzy Information and Decision Processes*; Gupta, M.M. , Sanchez, E., Eds; North-Holland: Amsterdam, Netherlands, 1982; pp. 167-182. <https://hal.science/hal-04072719v1/file/Dubois82.pdf> (accessed on 25 August 2024)
63. Dubois, D.; Prade, H. Unfair coins and necessity measures: towards a possibilistic interpretation of histograms. *Fuzzy Sets Syst.* **1983**, *10*, 15-20. [https://doi.org/10.1016/S0165-0114\(83\)80099-2](https://doi.org/10.1016/S0165-0114(83)80099-2)
64. Dubois, D.; Prade, H. Fuzzy sets and statistical data. *Eur. J. Oper. Res.* **1986**, *25*, 345-356. [http://doi.org/10.1016/0377-2217\(86\)90266-3](http://doi.org/10.1016/0377-2217(86)90266-3)
65. Yamada, K. Probability-possibility transformation based on evidence theory. In *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference* (vol.1), Vancouver, Canada, 25 July 2001, pp. 70-75. <http://doi.org/10.1109/NAFIPS.2001.944229>
66. Yager, R. G. Entropy and specificity in a mathematical theory of evidence. *Int. J. Gen. Syst.* **1983**, *9*(4), 249-260. <http://doi.org/10.1080/03081078308960825>
67. Jenhani, I.; Khlifi, G.; Sidiropoulos, P.; Jansen, H.; Frangou, G. Non-specificity-based supervised discretization for possibilistic classification. In *Scalable Uncertainty Management*; Dupin de Saint-Cyr, F., Öztürk-Escoffier, M., Potyka, N., Eds; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2022; Volume 13562 pp. 249-262 [https://doi.org/10.1007/978-3-031-18843-5\\_17](https://doi.org/10.1007/978-3-031-18843-5_17)
68. Klir, G.J.; Mariano, M. On the uniqueness of possibilistic measure of uncertainty and information. *Fuzzy Sets Syst.* **1987**, *24*(2), 197-219. [https://doi.org/10.1016/0165-0114\(87\)90090-X](https://doi.org/10.1016/0165-0114(87)90090-X)
69. Dubois, D.; Prade, H. A note on measures of specificity for fuzzy sets. *Bulletin pour les Sous-ensembles Flous et Leurs Applications* **1984**, *19*, 83-89. <https://hal.science/hal-04583421> (accessed on 20 August 2024)
70. Delgado, M.; Moral, S. On the concept of possibility-probability consistency. *Fuzzy Sets Syst.* **1987**, *21*(3), 311-318. [https://doi.org/10.1016/0165-0114\(87\)90132-1](https://doi.org/10.1016/0165-0114(87)90132-1)
71. Klir, G. J. A principle of uncertainty and information invariance. *Int. J. Gen. Syst.* **1990**, *17*(2-3), 249-275. <http://dx.doi.org/10.1080/03081079008935110>
72. Sudkamp, T. On probability-possibility transformations. *Fuzzy Sets Syst.* **1992**, *51*(1), 73-81. [https://doi.org/10.1016/0165-0114\(92\)90077-H](https://doi.org/10.1016/0165-0114(92)90077-H)
73. Salicone, S.; Jetti, H.V. A general mathematical approach based on the possibility theory for handling measurement results and all uncertainties. *Metrology* **2021**, *1*, 76–92. <https://doi.org/10.3390/metrology1020006>
74. Oussalah, M. On the probability/possibility transformations: a comparative analysis. *Int. J. Gen. Syst.* **2000**, *29*(5), 671-718. <http://doi.org/10.1080/03081070008960969>
75. Mei, W.; Li, G.; Xu, Y.; Shi, L. Probability-possibility transformation under perspective of random-fuzzy dual interpretation of unknown uncertainty. *TechRxiv* **2022** (preprint). <https://doi.org/10.36227/techrxiv.16896205.v2>
76. Jin, L.; Kalina, M.; Mesiar, R.; Borkotokey, S. Characterization of the possibility-probability transformations and some applications. *Inf. Sci.* **2019**, *477*, 281-290. <http://doi.org/10.1016/j.ins.2018.10.060>



77. Casquilho, J. P. On the weighted Gini-Simpson index: estimating feasible weights using the optimal point and discussing a link with possibility theory. *Soft Comput.* **2020**, *24*, 17187–17194. <https://doi.org/10.1007/s00500-020-05011-6>
78. Liang, S.; Peng, C.; Liao, Z.; Wang, Y. State space approximation for general fractional order dynamic systems. *Int. J. Syst. Sci.* **2014**, *45*(10), 2203–2212. <https://doi.org/10.1080/00207721.2013.766773>
79. Dubois, D.; Prade, H. Practical methods for constructing possibility distributions. *Int. J. Intell. Syst.* **2015**, *31*(3), 215–239. <https://doi.org/10.1002/int.21782>
80. De Luca, A.; Termini, S. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Inform. Control* **1972**, *20*(4), 301–312. [http://doi.org/10.1016/S0019-9958\(72\)90199-4](http://doi.org/10.1016/S0019-9958(72)90199-4)
81. Mei, W.; Liu, L.; Dong, J. The integrated sigma-max system and its application in target recognition. *Inf. Sci.* **2021**, *555*, 198–214. <https://doi.org/10.1016/j.ins.2020.12.054>
82. Yager, R. R. On the specificity of a possibility distribution. *Fuzzy Sets Syst.* **1992**, *50*, 279–292. [https://doi.org/10.1016/0165-0114\(92\)90226-T](https://doi.org/10.1016/0165-0114(92)90226-T)
83. Klir, G. J.; Parviz, B. Probability-possibility transformations; a comparison. *Int. J. Gen. Syst.* **1992**, *21*(3), 291–310. <http://dx.doi.org/10.1080/03081079208945083>
84. Voskoglou, M. G. Methods for assessing human-machine performance under fuzzy conditions. *Mathematics* **2019**, *7*, 230. <http://dx.doi.org/10.3390/math7030230>
85. Österreicher, F.; Casquilho, J. A. P. On the Gini-Simpson index and its generalisation – a historic note. *S. Afr. Stat. J.* **2018**, *52*(2), 129–137. <https://doi.org/10.37920/sasj.2018.52.2.2>
86. Mei, W. Probability/possibility systems for modeling of random/fuzzy information with parallelization consideration. *Int. J. Fuzzy Syst.* **2019**, *21*, 1975–1987. <https://doi.org/10.1007/s40815-019-00627-9>
87. Giakoumakis, S.; Papadopoulos, B. Novel transformation of unimodal (a)symmetric possibility distributions into probability distributions. *Fuzzy Sets Syst.* **2024**, *476*, 108790. <https://doi.org/10.1016/j.fss.2023.108790>
88. Andrés-Sánchez, J. Calculating insurance claim reserves with an intuitionistic fuzzy chain-ladder method. *Mathematics* **2024**, *12*, 845. <http://doi.org/10.3390/math12060845>
89. Parkash, O.; Singh, V.; Sharma, R. Weighted entropic and divergence models in probability spaces and their solicitations for influencing an imprecise distribution. In *Reliability Engineering for Industrial Processes*; Kapur, P.K., Pham, H., Singh, G., Kumar, V., Eds; Springer Series in Reliability Engineering, 2024, pp. 213–229. Springer, Cham, Switzerland [https://doi.org/10.1007/978-3-031-55048-5\\_15](https://doi.org/10.1007/978-3-031-55048-5_15)
90. Sánchez, L.; Costa, N.; Couso, I.; Strauss, O. Integrating imprecise data in generative models using interval-valued variational autoencoders. *Inf. Fusion* **2025**, *114*, 102659. <http://dx.doi.org/10.1016/j.inffus.2024.102659>
91. Bonacic, M.; López-Ospina, H.; Bravo, C.; Pérez, J. A fuzzy entropy approach for portfolio selection. *Mathematics* **2024**, *12*, 1921. <https://doi.org/10.3390/math12131921>
92. Zadeh, L. A. The information principle. *Inf. Sci.*, **2015**, *294*, 540–549. <http://dx.doi.org/10.1016/j.ins.2014.09.026>
93. Zadeh, L.A. Toward a restriction-centered theory of truth and meaning (RCT). In *Fifty Years of Fuzzy Logic and its Applications*; Tamir, D., Rishe, N., Kandel, A., Eds; Studies in Fuzziness and Soft Computing; Springer. Cham, Switzerland, 2015; Volume 326, pp. 1–24. [https://doi.org/10.1007/978-3-319-19683-1\\_1](https://doi.org/10.1007/978-3-319-19683-1_1)
94. Murari, A.; Rossi, R.; Spolladore, L.; Lungaroni, M.; Gaudio, P.; Gelfusa, M. A practical utility-based objective approach to model selection for regression in scientific applications. *Artif. Intell. Rev.* **2023**, *56*, S2825–S2859. <https://doi.org/10.1007/s10462-023-10591-4>
95. Miao, J.; Xing, H. Dynamic discrete choice under rational inattention. *Econ. Theory* **2024**, *77*, 597–652. <https://doi.org/10.1007/s00199-023-01512-1>
96. Casquilho, J. P. Discussing an expected utility and weighted entropy framework. *Nat. Science* **2014**, *6*, 545–551. <http://dx.doi.org/10.4236/ns.2014.67054>
97. Casquilho, J. P.; Rego, F. C. Discussing landscape compositional scenarios generated with maximization of non-expected utility decision models based on weighted entropies. *Entropy* **2017**, *19*, 66. <https://doi.org/10.3390/e19020066>

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