Probability-Possibility Transformation under Perspective of Random-fuzzy Dual Interpretation of Unknown Uncertainty

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

28-10-2021 / 01-11-2021

CITATION

Mei, wei; Li, Gan; Xu, Yan; Shi, Lin (2021): Probability-Possibility Transformation under Perspective of Random-fuzzy Dual Interpretation of Unknown Uncertainty. TechRxiv. Preprint.
<https://doi.org/10.36227/techrxiv.16896205.v1>

DOI

[10.36227/techrxiv.16896205.v1](https://doi.org/10.36227/techrxiv.16896205.v1)

Probability-Possibility Transformation under Perspective of Random-fuzzy Dual Interpretation of Unknown Uncertainty

Wei Mei, Gan Li, Yan Xu, Lin Shi

Abstract—Information-preservation is recognized as the only principle for probability-possibility transformation in this work and the normalized transformation is the right method. This is based on the viewpoint that the reason we can transfer probability and possibility is that we believe the uncertainty being handled can be information-equivalently described by both probability and possibility. That viewpoint is endorsed by the random-fuzzy dual interpretation of unknown uncertainty, which says that unknown uncertainty being estimated could be interpreted as either randomness or fuzziness, depending on the available prior information and the perspective of cognition and modeling. Information of uncertain variable is defined in this work as its distribution. The suggested information-preservation principle is different from Klir's principle, which is in fact an entropy-preservation principle. Then we investigated the problem of information preservation and propagation in parallel probability-possibility systems. By parallel, we mean the two uncertainty systems have the same priori information. After uncertainty propagation the two parallel systems will generally bifurcate, which means information preservation only holds locally between the two parallel systems. This observation accords with our intuition since probability and possibility use different normalizations as well as different disjunctive operators, which makes them two different uncertainty systems appropriate for randomness and fuzziness, respectively.

Index Terms—fuzzy uncertainty; probability theory; possibility theory; probability-possibility transformation; uncertainty inference; information of uncertainty variable.

I. INTRODUCTION

As an alternative method to fuzzy sets for handling fuzzy uncertainty, possibility theory has in recent decades received extensive attention [1-12, 44-46]. It was lately demonstrated that probability and possibility are customized for the measure of randomness and fuzziness, respectively [6]. As two kinds of widely acknowledged fundamental uncertainties of this world, randomness and fuzziness are recently defined in [6,8] as below:

Randomness is the occurrence uncertainty of the either-or outcome of a causal experiment, characterized by the lack of predictability in mutually exclusive outcomes.

Fuzziness is the classification uncertainty of the both-and outcome of a cognition experiment, characterized by the lack of clear boundary between non-exclusive outcomes.

Possibility theory has a structure that is parallel to probability theory and both methods describe resolution of uncertainty by

distribution of uncertain variable. Probability and possibility both measure (or score) the value of uncertainty with numbers ranging from 0 to 1, which makes up the primary similarity of probability and possibility. Nevertheless, the two concepts are in nature very different. Since random outcomes are *mutual exclusive*, probability requires that scores of every possible outcome of a random variable should sum to one, which is the key axiom of “additivity” [13-15]. In contrast, possibility is marked by the key axiom of “maxitivity” [1,2,6], which says that the possibility of two disjunctive *non-exclusive* fuzzy concepts is equal to the maximum of the two constituent possibilities. Possibility theory could be regarded as the foundation for fuzzy sets, since membership function of fuzzy sets could be recognized as likelihood function of possibility [16], and composition of fuzzy relations is equivalent to composition of conditional possibilities [5].

Despite their differences, many people are interested in connecting the two kinds of uncertainty measures of probability and possibility. Some of them try to interpret one uncertainty measure by another [17-20] whereas some others intend to build up conversion methods for them [21-25,42]. The motivation to study probability-possibility transformation has arisen not only from our desire to comprehend the relationship between the two theories of uncertainty, but also from some practical problems [21, 26]. Examples of these problems are: constructing a membership function of a fuzzy set from statistical data, constructing a probability measure from a given possibility measure in the context of decision making or systems modelling, or combining probabilistic and possibilistic information in expert systems [26-29].

By now various methods of probability-possibility transformation have been suggested in the literature. Except for the normalization requirements, they differ from one another substantially for the various principles that they are based upon [21, 23]. Among them, there are basically three transformation methods. The most common method is the normalized transformation, which is based on ratio scale to satisfy the normalization requirements of probability and possibility, respectively. Another method is the arising accumulation transformation [22, 24], which stresses that there are informational differences between possibility and probability measures, and transformation will lead to information loss or gain [22]. The third one is Klir's method, which adhibits uncertainty invariance principle, which claims that any

transformations that are not invariant with respect to the amount of uncertainty are ill-conceived [21].

This work is to reexamine the problem of probability-possibility transformation, but under the perspective of random-fuzzy dual interpretation of unknown uncertainty. For typical tasks of uncertain inference, such as estimation or recognition, the uncertainty handled is in fact unknown uncertainty that could be inferred or estimated by using available prior information and observations [7]. The dual interpretation says that the unknown uncertainty being handled could be interpreted as either randomness or fuzziness, depending on the available prior information and the perspective of cognition and modeling [7]. Though it does exist that in a certain situation one kind of interpretation is more reasonable than another kind of interpretation.

The dual interpretation suggests that the unknown uncertainty being handled could be simultaneously hence should be information-equivalently modeled as probability and possibility. That endorses the reason why we can transfer probability and possibility, which is that we believe uncertainty being handled in form of probability or possibility should have equivalent information. Otherwise, we will be upset with the conversion. In view of this, this work considers information preservation principle as the only principle for probability-possibility transformation, which will be elaborated in detail in section 4. Note that information preservation addressed in this work, is different from Klir's information preservation, which is entropy preservation in nature.

Be aware that uncertainty in form of probability or possibility has equivalent information does not mean the two systems of probability and possibility are equivalent. This work will as well investigate the problem of information preservation and propagation in parallel probability/possibility systems, which is section 5 and 6. By parallel, we mean the two uncertainty systems have the same priori conditions in term of equivalent information. We will see the two parallel systems will generally bifurcate after uncertainty propagation, which means information preservation only holds locally between the two parallel systems. The conclusion we have drawn are meaningful for understanding the two uncertainty systems, which are suitable for handling randomness and fuzziness, respectively.

The rest of the paper is organized as follows. Section 2 reviews fundamental aspects of possibility theory. Section 3 reviews three classic methods for probability-possibility transformation. Section 6 concludes the paper.

II. A REVIEW OF POSSIBILITY THEORY

2.1 Random/Fuzzy Variables

Definition 2.1. A *random variable* X is a variable whose value x_i is subject to variations due to random uncertainty. A random variable can take on a set of possible values in a random sample space Ω , or its generated event space $F \subseteq 2^\Omega$.

Definition 2.2. A *fuzzy variable* X is a variable whose value x_i is subject to variations due to fuzzy uncertainty. A fuzzy variable can take on a set of possible values in a fuzzy sample space Ψ , or its generated event space $\Sigma \subseteq 2^\Psi$.

Remark: A random variable should be modeled by probability

and fuzzy variable by possibility [4, 6]. Though events in $F \subseteq 2^\Omega$ and $\Sigma \subseteq 2^\Psi$ are both not mutually exclusive, the structures of random event space $F \subseteq 2^\Omega$ and fuzzy event space $\Sigma \subseteq 2^\Psi$ are not the same because their corresponding sample spaces Ω and Ψ are defined differently [6].

2.2 Unknown Uncertainty and Uncertain Variable

Definition 2.3. *Unknown uncertainty* is the uncertainty carried by an unknown quantity that is assumed to take on N possible outcomes $Y = \{x_1, x_2, \dots, x_N\}$.

Definition 2.4. *Unknown uncertain variable* (in short, *uncertain variable*) X is a variable whose value x_i is subject to variation due to unknown uncertainty. An uncertain variable can take on a set of possible values from $Y = \{x_1, x_2, \dots, x_N\}$, or its generated power set $\Xi \subseteq 2^Y$.

Remark: Uncertain variable can be modeled as either random variable or fuzzy variable depending on the available prior information [7].

2.3 Possibility and Conditional Possibility

Definition 2.5. *Possibility* (axiomatic definition) on universal set Ψ is defined as a mapping $\pi: 2^\Psi \rightarrow [0,1]$ such that [1,2],

Axiom 1. $\pi(\phi) = 0$ for empty set ϕ .

Axiom 2. $\pi(\Psi) = 1$.

Axiom 3. $\forall A, B \subseteq \Psi, \pi(A \cup B) = \max\{\pi(A), \pi(B)\}$.

By Axiom 2 and 3 we can derive

$$\pi(\Psi) = \max_{x_i} \pi(x_i) = 1. \quad (1)$$

Eq. (1) indicates that at least one of the elements of Ψ should be fully possible, i.e. $\exists x_i$, such that $\pi(x_i) = 1$. In contrast, probability obeys to the sigma normalization below:

$$\sum_{i=1}^N p(x_i) = 1. \quad (2)$$

Suppose $\pi(x_i y_j)$ is the joint possibility distribution of fuzzy variables X and Y , then conditional possibility $\pi(y_j | x_i)$ is defined as below [1,2,30]

$$\pi(x_i y_j) = \pi(y_j | x_i) \pi(x_i), \quad (3)$$

where

$$\pi(x_i) = \max_{y_j} \pi(x_i y_j), \quad \pi(y_j) = \max_{x_i} \pi(x_i y_j). \quad (4)$$

III. CLASSIC METHODS FOR PROBABILITY - POSSIBILITY TRANSFORMATION

Among a few, there are basically three transformation methods, which are proposed based on different transformation principles [21,23]. Let $X = \{x_1, x_2, \dots, x_n\}$, $p_i = p(x_i)$, and $\pi_i = \pi(x_i)$. Assume that elements of X are ordered in such a way that probability distributions, $P = (p_1, p_2, \dots, p_n)$ and possibility distributions, $\Pi = (\pi_1, \pi_2, \dots, \pi_n)$ are always nonincreasing sequences. That is, $p_i \geq p_{i+1}$, and $\pi_i \geq \pi_{i+1}$ for

all $i = 1, 2, \dots, n - 1$.

3.1 Normalized Transformation

The most common normalized transformation is based on the ratio scale as expressed by [21,23]

$$p_i = \frac{\pi_i}{\sum_{j=1}^n \pi_j} \quad (5)$$

$$\pi_i = \frac{p_i}{\max_j} p_j \quad (6)$$

3.2 Arising Accumulation Transformation

The arising accumulation transformation is an asymmetric transformation defined by the equations [22-24]

$$\pi_i = \sum_{j=i}^n p_j \quad (7)$$

$$p_i = \sum_{j=i}^n \frac{\pi_j - \pi_{j+1}}{\pi_j} \quad (8)$$

The transformation $P \rightarrow \Pi$ of (7) is based on the principle of maximum specificity, which aims at finding the most informative possibility distribution. A possibility distribution Π is said to be most informative if any other solution Π' is such that $\Pi \leq \Pi'$. The inverse transformation $\Pi \rightarrow P$ of (8) is based upon the principle of insufficient reason, which aims at finding the probability distribution that contains as much uncertainty as possible but that retains the features of possibility distribution [22]. The arising accumulation transformation satisfies two postulates below [22]:

1) Consistency condition: $p_i \leq \pi_i$ for all i . Here, the obtained possibility distribution should dominate the probability distribution.

2) Order preservation: $\pi_i \geq \pi_j$ if and only if $p_i \geq p_j$. Intuitively, if two worlds are ordered in a given way in P , then Π should preserve the same order.

3.3 Klir's Uncertainty-preservation Transformation

Transformation $P \leftrightarrow \Pi$ that preserve uncertainty were proposed by Klir [21]. It was shown by Geer and Klir that unique transformation of this kind exists only under log-interval scales as below [21]:

$$\pi_i = \left(\frac{p_i}{p_1} \right)^\alpha \quad (9)$$

$$p_i = \frac{\pi_i^{1/\alpha}}{\sum_{j=1}^n \pi_j^{1/\alpha}} \quad (10)$$

where the value of α is determined by solving (11) below

$$H(p) = N(\pi) + S(\pi) \quad (11)$$

which expresses the requirement that the amount of uncertainty be preserved when Π is transformed into P or vice versa. $H(p)$ is the well-known Shannon entropy of probability distribution given by [31]

$$H(p) = -\sum_{i=1}^n p_i \log_2 p_i \quad (12)$$

And $N(\pi)$ and $S(\pi)$ represent nonspecificity and strife of possibility distribution [21], respectively, given by

$$N(\pi) = \sum_{i=2}^n (\pi_i - \pi_{i+1}) \log_2 i \quad (13)$$

$$S(\pi) = \sum_{i=2}^n (\pi_i - \pi_{i+1}) \log_2 \frac{i}{\sum_{j=1}^i \pi_j} \quad (14)$$

It is known that Klir's transformations (11) and (12) satisfy the general possibility-probability consistency conditions: $p_i \leq \pi_i$ for all i . Note that given $\alpha = 1$, Klir's method will be the same as the normalized transformation.

IV. INFORMATION PRESERVATION IN PROBABILITY - POSSIBILITY TRANSFORMATION

4.1 The Dual Interpretation of Unknown Uncertainty

For the task of uncertain inference, interpretation of the uncertainty carried by the unknown quantity relies on the available prior information and the perspective of cognition and modeling [7]. For radar target recognition, e.g., the uncertainty of the unknown target type can be interpreted as either randomness or fuzziness. Under the perspective of occurrence of objective event, the unknown target type X should take place as an objective event x_i from the type set of Y that consists of possible but exclusive target types. In such a case, the unknown target type should be interpreted as randomness. Under the perspective of subjective cognition of concept, the unknown target type X could be simultaneously classified into more than one outcome from the type set of Y that consists of possible but non-exclusive target types, by using the available observation data and through the process of feature extraction. In such a case, the unknown target type should be interpreted as fuzziness. The fuzzy uncertainty arisen here has common in essence with that of the natural fuzzy concept such as Young and Senior, i.e., fuzziness is caused by the overlap of their intensions. Here intension indicates the internal content (abstracted feature) of a concept that constitutes its formal definition [6,43].

As discussed in the introduction, the dual interpretation of unknown uncertainty suggests that the unknown uncertainty being handled could be modeled as probability and possibility, simultaneously as well information-equivalently.

4.2 Information of Uncertain Variable

For a clear understanding of information, we have referred to both academic literature [31-36] and online dictionary [37,38]. Information is notoriously a polymorphic phenomenon and a polysemantic concept so, as an explicandum, it can be associated with several explanations, depending on the level of abstraction adopted and the cluster of requirements and desiderata orientating a theory [32]. Thus, Shannon [33] and Weaver [34] supported a tripartite analysis of information in terms of 1) technical problems concerning the quantification of information and dealt with by Shannon's theory; 2) semantic problems relating to meaning and truth; and 3) what he called "influential" problems concerning the impact and effectiveness of information on human behavior, which he thought had to

play an equally important role [32].

According to [32], the General Definition of (semantic) Information (GDI) has become an operational standard in the field of Information Sciences, which is clearly formulated as a tripartite definition (Fig. 1).

GDI) σ is an instance of information, understood as semantic content, if and only if:
 GDI.1) σ consists of n data (d), for $n \geq 1$;
 GDI.2) the data are *well-formed* (wfd);
 GDI.3) the wfd are *meaningful* (mwfd = δ).

Fig. 1. The General Definition of Information (GDI)

According to Wikipedia, information is the resolution of uncertainty [37]. Merriam-Webster gives a list of definitions for information, among which the following is related to the measure of uncertainty.

Information is a numerical quantity that measures the uncertainty in the outcome of an experiment to be performed [38].

This work agrees that information may manifest itself in different levels of abstraction. For instance, a word can carry the basic level of information, a sentence consisting of some words can carry more and higher level of information, and a paragraph (or even a paper) may contain much more comprehensive information. We can also say there are concept level of information, proposition level of information, and higher level of information. High level of information can be constructed by lower level of information. This work attributes information related to uncertain variable to concept-level information, and will next limit discussions on information of uncertain variable, whose definition is given below.

Definition 4.1. Information I of uncertain variable is the resolution of uncertainty, which is determined by the distribution of the uncertain variable X .

For random variable X with distribution $p(x_i)$, its random information is determined by

$$I(p) = \{I_p(x_i) | i = 1, 2, \dots, n\} = P \quad (15)$$

For fuzzy variable X with distribution $\pi(x_i)$, its fuzzy information is given by

$$I(\pi) = \{I_\pi(x_i) | i = 1, 2, \dots, n\} = \Pi \quad (16)$$

Remark: It is straightforward to verify that definition 4.1 of information of uncertain variable satisfies GDI. This definition of information is different from the concept of information entropy $H(p)$, which has been proposed for measuring information capacity. In communication, entropy represents the average rate at which information is produced by a stochastic source of data [37]. The definition of informationally equivalent is given below.

4.3 Definition to Informationally-equivalent

Definition 4.2. A random variable and a fuzzy variable are said to be *informationally equivalent* if their distributions satisfy the normalization requirement of (5) or (6).

Remark: The above definition to informationally-equivalent is natural since normalization of (5) or (6) is the minimum

requirement for a transformation between a possibility distribution and a probability distribution. It is safe to say such a normalization of proportional operation will not lead to information loss or gain. This point can be justified by the equivalence of uniform distribution of probability and possibility and the equivalence of the means (expected values) of probability/possibility functions. As we all know uniform distribution indicates innocent for both probability and possibility, and it can be easily verified that uniform distributions of probability and possibility satisfy normalizations of (5) and (6). As for their expected values, let us see the center of gravity (COG) of possibility function, which is defined as [39]

$$u_{\text{COG}} \triangleq \frac{\int_u u \cdot \pi(u|e) du}{\int_u \pi(u|e) du} \quad (17)$$

which is a concept parallel to the expected value of probability function. By using (5), it can be easily derived as in (18) that COG is equal to the mean of probability function. We therefore call u_{COG} as defined in (17) the mean (expected value) of possibility function.

$$u_{\text{COG}} = \int_u u \frac{\pi(u|e)}{\int_u \pi(u|e) du} du = \int_u u p(u|e) du \quad (18)$$

Eq. (18) reveals that probability/possibility functions with equivalent information have identical expected value and identical variance, as well. Given Gaussian possibility function as below [11,40]

$$\bar{N}(x; \mu, \Sigma) = \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \quad (19)$$

for some mean $\mu \in \mathbb{R}^{d_x}$ and for some covariance matrix $\Sigma \in \mathbb{R}^{d_x \times d_x}$, its counterpart of probability function $N(x; \mu, \Sigma)$ can be obtained by using the transformation of (5) or (6).

4.4 The Principle of Information-preservation

Definition 4.3. Uncertain variable X under transformation between probability and possibility should follow *the principle of information-preservation*, i.e., uncertain variable X in forms of probability and possibility should be informationally equivalent.

Remark: This principle is different from Klir's uncertainty-preservation principle, which is in fact an entropy-preservation principle. We conclude that the normalized transformation (5) and (6) is the right method for probability-possibility transformation, which is among all the simplest but graceful one.

4.5 A Transformation Example with Dual Interpretation of Unknown Uncertainty

Example 4.1. Let us consider an example of vote experiment similar to that in [8]. Suppose two groups of people, with each group having 100 persons, are called on. After showing a picture of a person, they are asked to answer whether he is strong, thin or between. Obviously, there is uncertainty involved in the answer. We tend to recognize this uncertainty as fuzzy uncertainty and prefer a possibility expression for the measure of the unknown quantity (physique) of being strong, thin or between. However, we will see below that the uncertainty can as well be recognized as random uncertainty

only if we carry out the experiment in a different way.

Group 1: Voters are allowed to vote 1~3 choices from the outcomes (strong, thin or between) simultaneously, which means this is a both-and choice. By definition, the uncertainty involved should be recognized as fuzziness and could be measured by possibility. And we got the vote results as in Table 1.

Group 2: Each voter is required to vote only one of the outcomes (strong, thin or between), which means this is an either-or choice. By definition, the uncertainty involved should be recognized as randomness and could be measured by probability. And we got the vote results as in Table 2.

As we can see in Table 1, strong has 100 votes whereas between has 30 votes, and thin has 0 vote. Therefore, the person is voted to be strong with possibility of 1, and to be between with possibility of 0.3. And these possibilities can be transformed into probabilities by using (5) or the method shown in Table 1. As we can see in Table 2, among 100 people, 80 people vote strong, and 20 people vote between and no one votes thin. Therefore, the person is voted to be strong with probability of 0.8, and to be between with probability of 0.2.

TABLE 1. VOTE RESULTS OF GROUP 1

	strong	between	thin
votes	100	30	0
possibility	100/100=1	30/100=0.3	0
transformed probability	100/130=0.77	30/130=0.23	0

TABLE 2. VOTE RESULTS OF GROUP 2

	strong	between	thin
votes	80	20	0
probability	80/100=0.8	20/100=0.2	0

From this vote experiment, we would like to ask two questions:

- 1) Can we do a transformation from possibility to probability?
- 2) If we can, what will be the principle guiding the transformation?

The answer to the first question is we can. We learned from the experiment that the uncertainty involved, which is fuzzy uncertainty in nature, can be recognized as random uncertainty as well only if the experiment is carried out in a different way. Therefore, the possibility for describing fuzziness can be transformed into probability to have a random sight-view on the fuzzy uncertainty. Since we are handling the same uncertainty carried by unknown quantity, the expressions for it before and after transformation should contain equivalent information. Otherwise, we will be upset with the conversion. In view of this, the answer to the second question is that information preservation principle (under the information viewpoint of uncertainty) should be the only principle for probability-possibility transformation.

V. INFORMATION PRESERVATION IN PARALLEL PROBABILITY/ POSSIBILITY SYSTEMS

We in this section investigate whether information preservation holds in parallel systems of probability and possibility. By parallel, we mean the two uncertainty systems have the same priori conditions in term of equivalent information.

5.1 Information Preservation in Event Space of Parallel Probability/ Possibility Systems

Proposition 5.1. Given informationally equivalent distributions $p(x_i)$ and $\pi(x_i)$ ($x_i \in \Omega$) in the sample space Ω , information preservation between distributions $p(A)$ and $\pi(A)$ ($A \in 2^\Omega$) in the event space 2^Ω does not hold.

Proof: This can be verified by a numeric example. Given informationally equivalent distributions $p(x_i)$ and $\pi(x_i)$ ($x_i \in \Omega$) as in Table 3, $p(A)$ and $\pi(A)$ ($A \in 2^\Omega$) in Table 4 are not informationally equivalent.

TABLE 3. $p(x_i) / \pi(x_i)$

x_i	x_1	x_2	x_3	x_4	x_5
p	0.5	0.2	0.1	0.1	0.1
π	1	0.4	0.2	0.2	0.2

TABLE 4. $p(A) / \pi(A)$

A	$x_1 \cup x_2$	$x_3 \cup x_4$	x_5
p	0.7	0.2	0.1
π	1	0.2	0.2

5.2 Information Preservation in 2-Ary Distribution of Parallel Probability/ Possibility Systems

We consider two situations as given in Propositions 5.2 and 5.3, respectively.

Proposition 5.2. Given informationally equivalent $p(x_i) / \pi(x_i)$ and $p(y_j|x_i) / \pi(y_j|x_i)$, information preservation in joint distributions $p(x_i y_j) / \pi(x_i y_j)$ does not hold.

Proof: We can verify by (20) that joint distribution $p(x_i y_j)$ do not satisfy (5).

$$p(x_i y_j) = p(x_i) p(y_j|x_i) = \frac{\pi(x_i)}{\sum_{x_k} \pi(x_k)} \cdot \frac{\pi(y_j|x_i)}{\sum_{y_l} \pi(y_l|x_i)} \neq \frac{\pi(x_i) \pi(y_j|x_i)}{\sum_{x_k} \sum_{y_l} \pi(x_k) \pi(y_l|x_k)} = \frac{\pi(x_i y_j)}{\sum_{x_k} \sum_{y_l} \pi(x_k y_l)} \quad (20)$$

where the expression to the right of the second equal-sign comes from the substitute of $p(x_i)$ and $p(y_j|x_i)$ by $\pi(x_i)$ and $\pi(y_j|x_i)$ using (5).

Similarly, we can verify by (21) that joint distribution $\pi(x_i y_j)$ do not satisfy (6).

$$\pi(x_i y_j) = \pi(x_i) \pi(y_j|x_i) = \frac{p(x_i)}{\max_{x_k} p(x_k)} \cdot \frac{p(y_j|x_i)}{\max_{y_l} p(y_l|x_i)} \neq \frac{p(x_i) p(y_j|x_i)}{\max_{x_k} \max_{y_l} p(x_k) p(y_l|x_k)} = \frac{p(x_i y_j)}{\max_{x_k} \max_{y_l} p(x_k y_l)} \quad (21)$$

Proposition 5.3. Given informationally equivalent joint distributions $p(x_i y_j) / \pi(x_i y_j)$, the derived conditional distributions $p(y_j|x_i) / \pi(y_j|x_i)$ hold information preservation; whereas the derived marginal distributions $p(x_i) / \pi(x_i)$ by projection do not hold information preservation.

Proof: Follow the line of (20) and (21), we can verify by (22) and (23) that the derived conditional distributions $p(y_j|x_i)$ and $\pi(y_j|x_i)$ do satisfy (5) and (6), respectively.

$$\begin{aligned}
p(y_j|x_i) &= \frac{p(x_i y_j)}{\sum_{y_l} p(x_i y_l)} = \frac{\frac{\pi(x_i y_j)}{\sum_{x_k y_l} \pi(x_k y_l)}}{\sum_{y_l} \left\{ \frac{\pi(x_i y_l)}{\sum_{x_k y_l} \pi(x_k y_l)} \right\}} \\
&= \frac{\pi(x_i y_j)}{\sum_{y_l} \pi(x_i y_l)} = \frac{\pi(x_i) \pi(y_j|x_i)}{\sum_{y_l} \pi(x_i) \pi(y_l|x_i)} = \frac{\pi(y_j|x_i)}{\sum_{y_l} \pi(y_l|x_i)} \quad (22)
\end{aligned}$$

$$\begin{aligned}
\pi(y_j|x_i) &= \frac{\pi(x_i y_j)}{\max_{y_l} \pi(x_i y_l)} = \frac{\frac{p(x_i y_j)}{\max_{y_l} p(x_k y_l)}}{\max_{y_l} \left\{ \frac{p(x_i y_l)}{\max_{y_l} p(x_k y_l)} \right\}} \\
&= \frac{p(x_i y_j)}{\max_{y_l} p(x_i y_l)} = \frac{p(x_i) p(y_j|x_i)}{\max_{y_l} p(x_i) p(y_l|x_i)} = \frac{p(y_j|x_i)}{\max_{y_l} p(y_l|x_i)} \quad (23)
\end{aligned}$$

Similarly, we can verify by (24) and (25) that the derived marginal distributions $p(x_i)$ and $\pi(x_i)$ do not satisfy (5) and (6).

$$\begin{aligned}
p(x_i) &= \sum_{y_j} p(x_i y_j) = \sum_{y_j} \frac{\pi(x_i y_j)}{\sum_{x_k y_l} \pi(x_k y_l)} \\
&\neq \max_{y_j} \frac{\pi(x_i y_j)}{\sum_{x_k y_l} \pi(x_k y_l)} \neq \frac{\pi(x_i)}{\sum_{x_k} \pi(x_k)} \quad (24)
\end{aligned}$$

$$\begin{aligned}
\pi(x_i) &= \max_{y_j} \pi(x_i y_j) = \max_{y_j} \frac{p(x_i y_j)}{\max_{x_k y_l} p(x_k y_l)} \\
&\neq \sum_{y_j} \frac{p(x_i y_j)}{\max_{x_k y_l} p(x_k y_l)} \neq \frac{p(x_i)}{\max_{x_k} p(x_k)} \quad (25)
\end{aligned}$$

People may worry that two inequalities in (24) and (25) may lead $p(x_i) / \pi(x_i)$ happen to be informationally equivalent. This worry can be eliminated by a proof by contradiction. Let us suppose $p(x_i) / \pi(x_i)$ are informationally equivalent, and by (22) and (23) we know $p(y_j|x_i) / \pi(y_j|x_i)$ are informationally equivalent, then according to Proposition 5.2, $p(x_i y_j) / \pi(x_i y_j)$ will not be informationally equivalent, which is contradicted to the prior setting of Proposition 5.3.

VI. INFORMATION PROPAGATION DOWN THROUGH PARALLEL PROBABILITY/ POSSIBILITY INFERENCES

We in this section compare information propagation down through parallel probability/ possibility inferences by two typical mechanisms: probability/ possibility update and combination of stochastic/ fuzzy relations.

6.1 Information Propagation in Uncertainty Update

Information propagation in form of probability update takes the form of (26) below,

$$p(x_i|y_j) = \frac{p(x_i) p(y_j|x_i)}{\sum_{x_k} p(x_k) p(y_j|x_k)} \quad (26)$$

where $p(x_i|y_j)$ is posteriori probability, $p(x_i)$ is priori probability, and $p(y_j|x_i)$ is probability likelihood of x_i .

Possibility update, as given below, has a parallel structure to Bayesian update [5,41]

$$\pi(x_i|y_j) = \frac{\pi(x_i) \pi(y_j|x_i)}{\pi(y_j)} = \frac{\pi(x_i) \pi(y_j|x_i)}{\max_{x_k} \pi(x_k) \pi(y_j|x_k)}, \quad (27)$$

where $\pi(x_i|y_j)$ is posteriori possibility, $\pi(x_i)$ is priori possibility, and $\pi(y_j|x_i)$ is possibility likelihood of x_i .

Proposition 6.1. Given informationally equivalent priori conditions $p(x_i) / \pi(x_i)$ and $p(y_j|x_i) / \pi(y_j|x_i)$, the updated conditional distributions $p(x_i|y_j) / \pi(x_i|y_j)$ are not informationally equivalent.

Proof: Substitute the equivalent $p(x_i)$ and $p(y_j|x_i)$ for $\pi(x_i)$ and $\pi(y_j|x_i)$ in (27), we have

$$\begin{aligned}
\pi(x_i|y_j) &= \frac{\pi(x_i) \pi(y_j|x_i)}{\max_{x_k} \pi(x_k) \pi(y_j|x_k)} \\
&= \frac{\frac{p(x_i)}{\max_{x_k} p(x_k)} \frac{p(y_j|x_i)}{\max_{y_l} p(y_l|x_i)}}{\max_{x_k} \left\{ \frac{p(x_k)}{\max_{x_k} p(x_k)} \frac{p(y_j|x_k)}{\max_{y_l} p(y_l|x_k)} \right\}} = \frac{\frac{p(x_i) p(y_j|x_i)}{\max_{y_l} p(y_l|x_i)}}{\max_{x_k} \left\{ \frac{p(x_k) p(y_j|x_k)}{\max_{y_l} p(y_l|x_k)} \right\}}. \quad (28)
\end{aligned}$$

Define $\hat{p}(x_i|y_j)$ that is equivalent to $\pi(x_i|y_j)$ as follows:

$$\hat{p}(x_i|y_j) \triangleq \frac{\pi(x_i|y_j)}{\sum_{x_k} \pi(x_k|y_j)} \quad (29)$$

Then by putting (28) into (29) and with the offset of the denominator of (28), we have

$$\hat{p}(x_i|y_j) = \frac{p(x_i) p(y_j|x_i) h(x_i)}{\sum_{x_k} p(x_k) p(y_j|x_k) h(x_k)} \neq p(x_i|y_j) \quad (30)$$

where

$$h(x_i) = \frac{1}{\max_{y_l} p(y_l|x_i)} \quad (31)$$

As we can see from (27) and (30), after uncertainty propagation, information preservation will not hold between the updated probability/possibility.

6.2 Information Propagation in Combination of Uncertainty Relations

Combination of stochastic relations. A “sigma-product” combination of random relations $p(x_i|z_k)$ and $p(z_k|y_j)$ derives a combined random relation $p(x_i|y_j)$, that is

$$p(x_i|y_j) = \sum_{z_k} p(x_i, z_k|y_j) = \sum_{z_k} p(x_i|z_k) p(z_k|y_j) \quad (32)$$

where x_i and y_j are assumed to be stochastically independent given z_k .

Combination of fuzzy relations. A “max-product” combination of fuzzy relations $\pi(x_i|z_k)$ and $\pi(z_k|y_j)$ derives a combined fuzzy relation $\pi(x_i|y_j)$, that is [4,5]

$$\pi(x_i|y_j) = \max_{z_k} \pi(x_i, z_k|y_j) = \max_{z_k} \pi(x_i|z_k) \pi(z_k|y_j) \quad (32)$$

where x_i and y_j are assumed to be possibilistically independent given z_k .

Proposition 6.2. Given informationally equivalent relations $p(x_i|z_k) / \pi(x_i|z_k)$ and $p(z_k|y_j) / \pi(z_k|y_j)$, the combined relation $p(x_i|y_j) / \pi(x_i|y_j)$ are not informationally equivalent.

Proof: Substitute the equivalent $p(x_i|z_k)$ and $p(z_k|y_j)$ for $\pi(x_i|z_k)$ and $\pi(z_k|y_j)$ in (33), we have

$$\begin{aligned}\pi(x_i|y_j) &= \max_{z_k} \pi(x_i|z_k) \pi(z_k|y_j) \\ &= \max_{z_k} \left\{ \frac{p(x_i|z_k)}{\max_{x_l} p(x_l|z_k)} \cdot \frac{p(z_k|y_j)}{\max_{z_m} p(z_m|y_j)} \right\} = \frac{\max_{z_k} \left\{ \frac{p(x_i|z_k)p(z_k|y_j)}{\max_{x_l} p(x_l|z_k)} \right\}}{\max_{z_m} p(z_m|y_j)} \quad (34)\end{aligned}$$

With the offset of the denominator of (34) in (35), we can derive $\hat{p}(x_i|y_j)$ equivalent to $\pi(x_i|y_j)$ as follows.

$$\begin{aligned}\hat{p}(x_i|y_j) &\triangleq \frac{\pi(x_i|y_j)}{\sum_{x_l} \pi(x_l|y_j)} = \frac{\max_{z_k} \left\{ \frac{p(x_i|z_k)p(z_k|y_j)}{\max_{x_l} p(x_l|z_k)} \right\}}{\sum_{x_l} \max_{z_k} \left\{ \frac{p(x_l|z_k)p(z_k|y_j)}{\max_{x_l} p(x_l|z_k)} \right\}} \\ &= \frac{\max_{z_k} p(x_i z_k | y_j) h(z_k)}{\sum_{x_l} \max_{z_k} p(x_l z_k | y_j) h(z_k)} \neq p(x_i|y_j) = \frac{\sum_{z_k} p(x_i z_k | y_j)}{\sum_{x_l} \sum_{z_k} p(x_l z_k | y_j)} \quad (35)\end{aligned}$$

where

$$h(z_k) = \frac{1}{\max_{x_l} p(x_l|z_k)}. \quad (36)$$

As we can see from (35), after uncertainty propagation, information preservation will not hold between the combined statistic relation and the combined fuzzy relation.

VII. CONCLUSION

Under the perspective of the dual interpretation of unknown uncertainty, the normalized transformation is recognized as the right method for probability-possibility transformation, which satisfies the information-preservation principle. Given the same priori information, the two parallel systems of probability and possibility will generally bifurcate after uncertainty propagation. This observation accords with our intuition since probability and possibility use different normalizations as well as different disjunctive operators. As typical applications to target recognition and multiple-model state estimation, possibility inference exhibits significant better performance than the probability inference [4,6]; whereas the possibilistic stochastic filter performs identically as the standard Bayes filter [47], and would reduce to the same form as the Kalman filter in the linear-Gaussian case [48].

Future research will conduct more performance analysis on the difference of the two uncertainty systems of probability and possibility, and explore more applications including estimation & recognition in the framework of possibility theory.

REFERENCES

- [1] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* 1 (1978): 3–28. (Reprinted in *Fuzzy Sets and Systems* 100 (Supplement) (1999): 9–34)
- [2] B. Solaiman, É. Bossé, Possibility Theory for the Design of Information Fusion Systems, Springer, Cham, 2019
- [3] H. Yin, J. Huang and Y. -H. Chen, Possibility-Based Robust Control for Fuzzy Mechanical Systems, *IEEE Transactions on Fuzzy Systems*, doi: 10.1109/TFUZZ.2020.3028940.
- [4] W. Mei, L.M. Liu, J. Dong, The Integrated Sigma-Max System and Its Application in Target Recognition, *Information Sciences*, 555, May 2021: 198–214, DOI: 10.1016/j.ins.2020.12.054
- [5] W. Mei, Formalization of Fuzzy Control in Possibility Theory via Rule Extraction, *IEEE Access*, July 2019, DOI: 10.1109/ACCESS.2019.2928137
- [6] W. Mei, M. Li, Y.Z. Cheng and L.M. Liu, Sigma-max system induced from randomness and fuzziness, arXiv preprint arXiv:2110.07722 [cs.AI] (2021), <https://arxiv.org/abs/2110.07722>.
- [7] W. Mei, Y.F. Xu, L.M. Liu, Estimation & Recognition under Perspective of Random-Fuzzy Dual Interpretation of Unknown Quantity: with Demonstration of IMM Filter, arXiv preprint arXiv:2110.10572 [eess.SY] (2021), <http://arxiv.org/abs/2110.10572>.
- [8] W. Mei, Probability/possibility systems for modeling of random/fuzzy information with parallelization consideration, *International Journal of Fuzzy Systems*, 2019, DOI: 10.1007/s40815-019-00627-9
- [9] B. Alsahwa, B. Solaiman, S. Almouahed, et al. Iterative Refinement of Possibility Distributions by Learning for Pixel-Based Classification, *IEEE Transactions on Image Processing*, 2016, 25(8):3533–3545.
- [10] J. Houssineau, A Linear Algorithm for Multi-Target Tracking in the Context of Possibility Theory, *IEEE Transactions on Signal Processing*, vol. 69, pp. 2740–2751, 2021, doi: 10.1109/TSP.2021.3077304.
- [11] B. Ristic, J. Houssineau, S. Arulampalam, Target tracking in the framework of possibility theory: The possibilistic Bernoulli filter. *Information Fusion*, 2020, 62.
- [12] Ferrero, H. V. Jetti and S. Salicone, The Possibilistic Kalman Filter: Definition and Comparison with the Available Methods, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
- [13] I. Hacking, *The Emergence of Probability* (2nd ed). New York: Cambridge University Press (2006).
- [14] E.T. Jaynes, *Probability Theory: The Logic of Science*, edited by L. Bretthorst. Cambridge University Press, (2003).
- [15] B. Liu, *Uncertainty Theory*, 5rd ed., Springer-Verlag Berlin, 2015, <http://orsc.edu.cn/liu/ut.pdf>: 471–472
- [16] G. Coletti, D. Petturiti, B. Vantaggi, Fuzzy memberships as likelihood functions in a possibilistic framework, *International Journal of Approximate Reasoning*, 88, January 2017: 547–566
- [17] G. Coletti, R. Scozzafava, Conditional probability, fuzzy sets, and possibility: a unifying view, *Fuzzy Sets and Systems* 144 (2004): 227–249.
- [18] D. Dubois, S. Moral and H. Prade, A semantics for possibility theory based on likelihoods, *Journal of Mathematic Analysis and Applications*, 205 (1997): 359–380.
- [19] W. Mei, Bridging probability and possibility via Bayesian theorem, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 4 (2014): 615–626
- [20] Marco E.G.V. Cattaneo, The likelihood interpretation as the foundation of fuzzy set theory, *International Journal of Approximate Reasoning*, November 2017: 333–340
- [21] G. J. Klir, B. Parviz, Probability-possibility transformation: A Comparison, *International Journal of General Systems*, 21(3), 1992: 291–310
- [22] D. Dubois, L. Foulloy et al., Probability-possibility transformation, triangular fuzzy sets, and probabilistic inequalities, *Reliable Computing*, 10 (2004): 273–297.
- [23] M. S. Mouchaweh and P. Billaudel, Variable probability-possibility transformation for the diagnosis by pattern recognition, *Int. J. Computational Intelligence: Theory and Practice* 1 (2006): 9–21.
- [24] L. Jin, M. Kalina, R. Mesiar, Characterizations of the possibility-probability transformations and some applications, *Information Sciences* 477 (2019): 281 – 290
- [25] M. Oussalah, On the probability/possibility transformations: a comparative analysis, *Journal of General Systems*, 29(5) (2000): 671–718.
- [26] M. Pota, M. Esposito, and G. De Pietro, Transforming probability distributions into membership functions of fuzzy classes: A hypothesis test approach, *Fuzzy Sets and Systems*, 233 (2013): 52–73.
- [27] Ü. S. Sakallı, Ö. F. Baykoç, Can the uncertainty in brass casting blending problem be managed? A probability/possibility transformation approach, *Computers & Industrial Engineering*, 2011, 61(3):729–738.
- [28] J. L. Caillec, A. Itani, et al, Stock picking by Probability-Possibility approaches, *IEEE Transactions on Fuzzy Systems*, 25(2) (2017): 333–349.
- [29] R.R. Yager, Conditional Approach to possibility-probability fusion, *IEEE Transactions on Fuzzy Systems*, 1 (2012): 46–56
- [30] B. De Baets, E. Tsiporkova and R. Mesiar Conditioning in possibility with strict order norms, *Fuzzy Sets and Systems*, 106, 221–229, 1999.

- [31] C.E. Shannon, A Mathematical Theory of Communication, *Bell System Technical Journal*. 27 (3), July 1948: 379-423.
- [32] L. Floridi (2009), Philosophical Conceptions of Information. In: Sommaruga G. (eds) Formal Theories of Information. Lecture Notes in Computer Science, vol 5363. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-00659-3_2
- [33] C.E. Shannon, Collected Papers. In: Sloane, N.J.A., Wyner, A.D. (eds.). IEEE Press, New York (1993)
- [34] W. Weaver, The Mathematics of Communication. *Scientific American* 181(1), 11-15 (1949)
- [35] Mingers J, Standing C. What is Information? Toward a Theory of Information as Objective and Veridical. *Journal of Information Technology*. 2018;33(2):85-104. doi:10.1057/s41265-017-0038-6
- [36] Krzanowski R. What Is Physical Information? *Philosophies*. 2020; 5(2):10. <https://doi.org/10.3390/philosophies5020010>
- [37] https://en.wikipedia.org/wiki/Information#cite_ref-1 (2019). Accessed 8 Mar 2019
- [38] <https://www.merriam-webster.com/dictionary/information> (2019). Accessed 8 Mar 2019
- [39] H. J. Zimmermann, Fuzzy set theory – and its applications (4th ed.), Springer Science+Business Media New York, 2001.
- [40] J. Houssineau and A.N. Bishop. Smoothing and filtering with a class of outer measures. *SIAM Journal on Uncertainty Quantification*, 2018.
- [41] S. Lapointe, B. Bobée, Revision of possibility distributions: A Bayesian inference pattern, *Fuzzy sets and systems*, 116 (2) (2000): 119-140.
- [42] S. Benferhat, A. Levray, K. Tabia, On the Analysis of Probability-Possibility Transformations: Changing Operations and Graphical Models. In: Destercke S., Denoeux T. (eds) Symbolic and Quantitative Approaches to Reasoning with Uncertainty. ECSQARU 2015. Lecture Notes in Computer Science, vol 9161. Springer, Cham. https://doi.org/10.1007/978-3-319-20807-7_25
- [43] F. M. Djalal, J. A. Hampton, G. Storms, & T. Heyman, Relationship between Extensions and Intensions in Categorization: A Match Made in Heaven? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, October 2017
- [44] M. B. Ferraro, P. Giordani, Possibilistic and fuzzy clustering methods for robust analysis of non-precise data, *International Journal of Approximate Reasoning*, September 2017: 23-38
- [45] Z. He, L. Yu, Possibility distribution based lossless coding and its optimization, *Signal Processing*, Vol. 150, (2018): 122-134
- [46] Z. Wu, J. Xu, X. Jiang, et al, Two MAGDM models based on hesitant fuzzy linguistic term sets with possibility distributions: VIKOR and TOPSIS, *Information Sciences*, Vol. 473, (2019): 101-120
- [47] B. Ristic, J. Houssineau, and S. Arulampalam. Robust target motion analysis using the possibility particle filter. *IET Radar Sonar Navigation*, 13(1), 2019.
- [48] J. Houssineau and A.N. Bishop. Smoothing and filtering with a class of outer measures. *SIAM Journal on Uncertainty Quantification*, 2018.