

On the Analysis of Probability-Possibility Transformations: Changing Operations and Graphical Models

Salem Benferhat, Amélie Levray, and Karim Tabia

Univ Lille Nord de France, F-59000 Lille, France
UArtois, CRL UMR CNRS 8188, F-62300 Lens, France

Abstract. Representing and reasoning with uncertain information is a common topic in Artificial Intelligence. In this paper, we focus on probability-possibility transformations in the context of changing operations and graphical models. Existing works mainly propose probability-possibility transformations satisfying some desirable properties. Regarding the analysis of the behavior of these transformations with respect to changing operations (such as conditioning and marginalization), only few works addressed such issues. This paper concerns the commutativity of transformations with respect to some reasoning tasks such as marginalization and conditioning. Another crucial issue addressed in this paper is the one of probability-possibility transformations in the context of graphical models, especially the independence of events and variables.

Keywords: Probability-possibility transformations, marginalization, conditioning, graphical models

1 Introduction

Several frameworks exist for representing and reasoning with uncertain information. Probability and possibility theories are among the most commonly used. Probability theory is the oldest theory dealing with uncertainty and frequentist setting. The early works involving probability and possibility theories were devoted to establishing connections between these two frameworks (as in [19,12,15]). These works are mostly interested in finding desirable properties to satisfy and then proposing transformations that guarantee these properties. An example of such desirable properties is the consistency principle used to preserve as much information as possible.

Probability-possibility transformations are useful in many ways. For instance, an example of propagating probabilistic (stochastic) and possibilistic information in risk analysis is provided in [1]. Another motivation is the fact that probabilities are more suitable in a frequentist setting, but this requires a large number of data, and when data is not available in sufficient quantities then the possibilistic setting can fill this lack as in [13]. Another motivation for probability-possibility transformations is to use existing tools (e.g. algorithms and software) developed

in one setting rather than developing everything from scratch.

In this paper, we deal with probability-possibility transformations with respect to reasoning tasks and graphical models. On that matter, a few works have been published. In [18], the author address the commutativity of probability-possibility transformations with respect to some reasoning tasks. The authors in [16] study some issues related to transforming Bayesian networks into possibilistic networks. In [5], the authors deal with transforming probability intervals into other uncertainty settings. Note that in this paper, we are only interested in transformations from probability distributions into possibility distributions. Given a distribution encoding some uncertain information, be it possibilistic or probabilistic, we are supposed to be able to reason about events of interest. In this work, we are interested in studying complementary issues such as preserving marginalization, conditioning and independence relations. We analyze these issues when the available information is encoded by means of distributions or in a compact way in the form of graphical models. We show that there is no transformation from the probabilistic into the possibilistic setting that guarantee most of the reasoning tasks dealt with in this work. For instance, regarding preserving marginalization, we show that no transformation can preserve the relative order of arbitrary events even if it preserves the relative order of interpretations. When transforming probabilistic graphical models, the order of interpretations cannot be preserved neither. Before presenting our results, let us first recall some concepts and present some existing probability-possibility transformations.

2 A Refresher on probability and possibility theories and graphical models

Probability theory is a well-known and widely used uncertainty framework. One of the building blocks of this setting is the one of probability distribution p assigning a probability degree to each elementary state of the world. Probability theory is ruled by Kolmogorov's axioms (non negativity, normalization and additivity) and usually have two main interpretations (namely, the frequentist and subjective interpretations). Among the alternative uncertainty theories, possibility theory [19,8] is a well-known one. It is based on the notion of possibility distribution π which maps every state ω_i of the world Ω (the universe of discourse) to a degree in the interval $[0, 1]$ expressing a partial knowledge over the world. By convention, $\pi(\omega_i)=1$ expresses that ω_i is totally possible, while $\pi(\omega_i)=0$ means that this world is impossible. Note that possibility degrees are interpreted either i) *qualitatively* (in min-based possibility theory) where only the "ordering" of the values is important, or *quantitatively* (in product-based possibility theory) where the possibilistic scale $[0, 1]$ is quantitative as in probability theory. One of the main difference between probability theory and possibility theory is that the former is additive while the latter is maxitive ($\Pi(\phi \cup \psi) = \max(\Pi(\phi), \Pi(\psi)) \forall \phi, \psi \subseteq \Omega$).

Conditioning is an important belief change operation concerned with updating the current beliefs encoded by a probability or a possibility distribution

when a completely sure event (evidence) is observed. While there are several similarities between the quantitative possibilistic and the probabilistic frameworks (conditioning is defined in the same way following the so-called Dempster rule of conditioning), the qualitative one is significantly different. Note that the two definitions of possibilistic conditioning satisfy the condition: $\forall \omega \in \phi, \pi(\omega) = \pi(\omega|\phi) \otimes \Pi(\phi)$ where \otimes is either the product or min-based operator. In the quantitative setting, the product-based conditioning is defined as follows:

$$\pi(w_i|_p\phi) = \begin{cases} \frac{\pi(w_i)}{\Pi(\phi)} & \text{if } w_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Conditioning in the qualitative setting is defined as follows [11]:

$$\pi(w_i|_m\phi) = \begin{cases} 1 & \text{if } \pi(w_i) = \Pi(\phi) \text{ and } w_i \in \phi; \\ \pi(w_i) & \text{if } \pi(w_i) < \Pi(\phi) \text{ and } w_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Working directly with uncertainty (probability or possibility) distributions is not convenient in terms of spatial and temporal complexity. Indeed, the distribution size can become too large to be stored and manipulated. This is why belief graphical models [4] have been developed. They represent uncertain information in a more compact way, and multiple tools have been developed for inference.

Bayesian networks A Bayesian network [4] is specified by:

- A *graphical component* with vertices and edges forming a directed acyclic graph (DAG). Each vertex represents a variable A_i of the modeled problem and the edges encode independence relationships among variables.
- A *quantitative component*, where every variable A_i is associated with a local probability distribution $p(A_i|par(A_i))$ for A_i in the context of its parents, denoted $par(A_i)$.

A Bayesian network encodes a joint probability distribution using the following chain rule:

$$P(A_1, \dots, A_n) = \prod_{i=1}^n P(A_i|par(A_i)) \quad (3)$$

Bayesian networks are not only used to represent information but also to reason with it. Many algorithms for exact and approximate inferences exist for probabilistic graphical models [4].

Possibilistic networks A possibilistic network [3] is also specified by a graphical and a numeric component where the local tables are possibility distributions. The chain rule is defined as follows:

$$\pi(A_1, \dots, A_n) = \otimes_{i=1..n} \pi(A_i | par(A_i)) \quad (4)$$

where \otimes is either the product or min-based operator (namely, $\otimes = \min$ or $\otimes = *$). Unless otherwise stated, all that follows is valid in both the quantitative or qualitative possibilistic settings.

3 Probability-possibility transformations

In this section, we first review the main principles of probability-possibility transformations. In particular, since probability and possibility theories represent different kinds of uncertainty, there is a need to focus on the concept of consistency coined by Zadeh [19] and redefined by many authors like Dubois and Prade [7].

3.1 Basic principles for probability-possibility transformations

The first principle that transformations tried to satisfy is due to Zadeh [19]:

Zadeh consistency principle. Zadeh [19] measures the consistency between a probability and possibility distribution as follows:

$$C_z(\pi, p) = \sum_{i=1..n} \pi(\omega_i) * p(\omega_i). \quad (5)$$

where p and π are a probability and a possibility distributions respectively over a set of n worlds. It intuitively captures the fact that “*A high degree of possibility does not imply a high degree of probability, and a low degree of probability does not imply a low degree of possibility*”. The computed consistency degree is questionable [7,12] in the sense that two resulted possibility distributions can have the same consistency degree but do not contain the same amount of information.

Dubois and Prade consistency principle. Dubois and Prade [7] defined three postulates allowing to define the optimal transformation [7] which always exist and it is unique.

- *Consistency condition* states that for each event (ie. a set of worlds) $\phi \subseteq \Omega$, $P(\phi) \leq \Pi(\phi)$. Here, the obtained possibility distribution should dominate the probability distribution.
- *Preference preservation*: $\forall (\omega_1, \omega_2) \in \Omega^2$, $p(\omega_1) \geq p(\omega_2)$ iff $\pi(\omega_1) \geq \pi(\omega_2)$. Intuitively, if two worlds are ordered in a given way in p , then π should preserve the same order.
- *Maximum specificity principle*: This principle requires to search for the most specific possibility distribution that satisfies the two above conditions. Let π_1 and π_2 be two possibility distributions, π_1 is said to be more specific than π_2 if $\forall \omega_i \in \Omega$, $\pi_1(\omega_i) \leq \pi_2(\omega_i)$.

3.2 Transformation rules

Many probability-possibility transformations have been proposed in the literature. We cite the *Optimal transformation* (OT) [7], *Klir transformation* (KT) [12], *Symmetric transformation* (ST) [10], and *Variable transformation* (VT) [14]. The optimal transformation (OT) guarantees the most specific possibility distribution that satisfies Dubois and Prade’s consistency principle. It is defined as follows:

$$\pi(\omega_i) = \sum_{j/p(\omega_j) \leq p(\omega_i)} p(\omega_j). \quad (6)$$

Note that there exist transformations from the possibilistic setting into the probabilistic one [10] and into other uncertainty frameworks [5].

4 Transformations and changing operations

Our purpose in this paper is to study the commutativity of transformations on reasoning tasks. In [18], the author was the first to study this question but his focus was only if the resulted distributions are *identical*. He showed that there is no transformation satisfying commutativity of transformations with respect to operations like conditioning and marginalization. We use $\triangleright(p)=\pi$ to denote the transformation from a probability distribution into possibility distribution satisfying Dubois and Prade preference preservation principle. In the following, we study the commutativity of transformations with respect to i) the order of arbitrary events and ii) two changing operators that are marginalization and conditioning. We focus on these two issues especially for useful practical uses of transformations. In fact, among the most used queries in probabilistic models, we find MPE queries (searching for the most plausible explanations) and MAP (where given some observations, the objective is to find the most plausible values of some variables of interest) [4]. For instance, let $p(ABC)$ be a probability distribution over three binary variables A, B and C . Let $C=0$ be an observation. MPE query would be “which is the most probable interpretation for $p(A, B, C=0)$ ”. MAP query would be “which is the most probable set of interpretations for $p(A, B|C=0)$ ”. To answer such queries using probability-possibility transformations, it is necessary to study the commutativity of transformations with respect to the marginalization and conditioning operations.

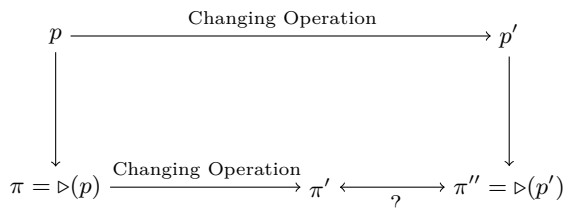


Fig. 1: Commutativity of operations

We consider operations on distributions as depicted on Figure 1. On one hand we obtain a possibility distribution by first applying an operation then the transformation, and on the other hand we obtain the possibility distribution by first transforming the probability distribution then applying the corresponding operation in the possibilistic setting. Our objective is to compare these distributions and see if they encode the same order.

We first consider the operation of marginalization which consists in building the marginal distributions from a joint distribution.

4.1 Marginalization and Transformations: Preservation of the order of arbitrary events

As said in the previous section, one of the principles of Dubois and Prade requires that the order of interpretations must be preserved, but nothing is said regarding arbitrary events (sets of interpretations). For instance, is it enough for

a transformation to preserve the order of interpretations to preserve the order of arbitrary events? Proposition 1 states that there is no probability-possibility transformation preserving the order of events.

Proposition 1. *Let \triangleright be a probability-possibility transformation operation (or function)¹. Then there exists a probability distribution p , $\phi \subseteq \Omega$, $\psi \subseteq \Omega$, with $\phi \neq \psi$, and $\pi = \triangleright(p)$ such that*

$$P(\phi) < P(\psi) \text{ holds but } \Pi(\phi) < \Pi(\psi) \text{ does not hold.}$$

The reason of loosing the strict order is due to the difference in behavior of the additivity axiom in the probabilistic setting and the maxitivity axiom of the possibilistic setting. As a consequence of Proposition 1, if the universe of discourse Ω is a cartesian product of a set of variable domains, then the marginalization over variables will not preserve the relative order of events after the transformation operation.

4.2 Conditioning and Transformations: Preservation of the order of arbitrary events

The question here is “*is the order of interpretations and arbitrary events preserved if we apply conditioning before or after transformation?*”.

Proposition 2 states that the order of elementary interpretations after conditioning is preserved if the used transformation preserves the order of interpretations.

Proposition 2. *Let $\phi \subseteq \Omega$ be an evidence. Let \triangleright be a probability-possibility transformation, p' be a probability distribution obtained by conditioning p by ϕ , $\pi'' = \triangleright(p')$ and π' is the possibility distribution obtained by conditioning $\pi = \triangleright(p)$ by ϕ . Then, $\forall \omega_i, \omega_j \in \Omega$, $\pi'(\omega_i) < \pi'(\omega_j)$ iff $\pi''(\omega_i) < \pi''(\omega_j)$.*

Proposition 2 is valid using both the product or min-based conditioning.

As a consequence of Proposition 2, if one is interested in MPE queries, then the answers of such queries are exactly the same if we condition then transform or first transform then condition. However, because of the loss of the order of events when marginalizing (see Proposition 1), then the answers to MAP queries will not be the same.

4.3 Independence relations and Transformations

When dealing with uncertain and incomplete information, the notion of independence² is very important. This subsection checks if the independence relation between events is preserved. Of course, the concept of independence is linked to the one of conditioning and marginalization. Proposition 3 states that there is no transformation operation \triangleright that preserves the independence relations.

¹ \triangleright is always assumed to satisfy Dubois and Prade consistency and preference preservation principle.

² Let α , ϕ and ψ be three arbitrary events, in probability theory (resp. possibility theory), ϕ is said to be independent of ψ in the context of α iff $P(\phi|\psi, \alpha) = P(\phi|\alpha)$ (resp. $\Pi(\phi|\psi, \alpha) = \Pi(\phi|\alpha)$).

Proposition 3. Let ϕ , ψ and $\alpha \subseteq \Omega$ be three events. Let \triangleright be a probability-possibility transformation operation. Then there exist a probability distribution p and $\pi = \triangleright(p)$ such that

$$P(\phi | \psi\alpha) = P(\phi | \alpha) \text{ but } \Pi(\phi |_{\otimes} \psi\alpha) \neq \Pi(\phi |_{\otimes} \alpha)$$

In Proposition 3, $|_{\otimes}$ denotes either the product or min-based conditioning operator. As a consequence, we can state that the independence of variables is not preserved either. This represents a major issue especially if one applies transformations to graphical models which are based on the concept of conditional independence relations.

5 Graphical models and Transformations

Let us first define a transformation of a probabilistic graphical model into a possibilistic one. We transform a Bayesian network into a possibilistic network as follows (as in [16]):

Definition 1. Let \mathcal{BN} be a Bayesian network over a set of variables $A = \{A_1, \dots, A_n\}$, \mathcal{PN} be a possibilistic network over the same set of variables A . \mathcal{PN} is obtained by a transformation operation \triangleright defined as follows:

- The graphical component of \mathcal{PN} is the same graph as the one of the Bayesian network \mathcal{BN} .
- The numerical component of \mathcal{PN} is such that every local probability table $p(A_i | \text{par}(A_i))$ is transformed with \triangleright into $\pi(A_i | \text{par}(A_i)) = \triangleright(p(A_i | \text{par}(A_i)))$.

The advantage of transforming a graphical model using Definition 1 is preserving independence relations, while computationally it is less consuming to transform a set of local tables than a whole joint distribution. The problem now is that there is no guarantee that the order of interpretations and events is preserved in the obtained possibilistic network and its underlying joint distribution. Figure 2 illustrates the issue of transforming a Bayesian network into a possibilistic one.

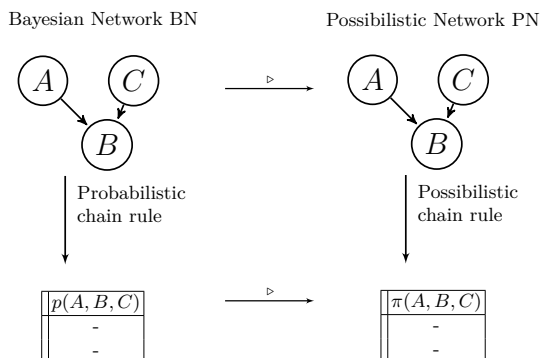


Fig. 2: Belief graphical model transformation

Let us now check if the order of interpretations induced by $p_{\mathcal{BN}}$ (the joint distribution encoded by the Bayesian network \mathcal{BN}) is preserved in the obtained joint possibility distribution $\pi_{\mathcal{PN}}$ (the joint distribution encoded by the possibilistic network \mathcal{PN}). Proposition 4 answers this question.


Proposition 4. *Let \triangleright be a probability-possibility transformation. Then there exist a Bayesian network \mathcal{BN} , $\omega_1 \in \Omega$ and $\omega_2 \in \Omega$ where:*


$$\pi'(\omega_1) < \pi'(\omega_2) \text{ does not imply } \pi''(\omega_1) < \pi''(\omega_2)$$

where: i) $\pi'(\omega) = \triangleright(p(\omega))$ and p is the joint distribution induced by \mathcal{BN} and ii) π'' is the joint distribution induced by \mathcal{PN} using Definition 1.

Example 1. Let \mathcal{BN} be the Bayesian network of Figure 3 over two disconnected variables A and B . Note that the probability distribution $p(A)$ in \mathcal{BN} is a permutation³ of the probability distribution $p(B)$. Hence, the transformation of $p(A)$ and $p(B)$ by \triangleright gives $\pi(A)$ and $\pi(B)$ where $\pi(B)$ is also a permutation of $\pi(A)$. In this example, since \triangleright is assumed to preserve the order of interpretations, we have $1 > \alpha_1 > \alpha_2 > \alpha_3$. The

A	$p(A)$	$\pi(A)$
a_1	0.4	1
a_2	0.2	α_2
a_3	0.25	α_1
a_4	0.15	α_3





B	$p(B)$	$\pi(B)$
b_1	0.15	α_3
b_2	0.2	α_2
b_3	0.25	α_1
b_4	0.4	1

Fig. 3: Example of Bayesian-possibilistic network transformation.

probability and possibility degrees of interpretations a_1b_1 and a_2b_2 are

$$\begin{aligned} - p(a_1b_1) &= 0.4 * 0.15 = 0.06 \\ - p(a_2b_2) &= 0.2 * 0.2 = 0.04 \end{aligned} \quad \text{then, } p(a_1b_1) > p(a_2b_2) \quad (a)$$

$$\begin{aligned} - \pi(a_1b_1) &= \alpha_3 \\ - \pi(a_2b_2) &= \alpha_2 \end{aligned} \quad \text{then, } \pi(a_1b_1) < \pi(a_2b_2) \quad (b)$$

From (a) and (b) one can see that the relative order of interpretations is reversed whatever is the used transformation in the ordinal setting. In the same way, in the quantitative setting, the relative order of interpretations can not be preserved by any transformation.

6 Related works and discussions

This paper dealt with some issues about probability-possibility transformations especially those regarding reasoning tasks and graphical models. We showed that there is no transformation that can preserve the order of arbitrary events through some reasoning operations like marginalization. As for the independence of events and variables, we showed that there is no transformation that preserves the independence relations. When the uncertain information is encoded by means

³ The permutation property of probability-possibility transformations is discussed in [18].

of graphical models, we showed that no transformation can preserve the order of interpretations and events.

In the literature, there are two works in particular that dealt with the issues of our work. First, in [16] the authors studied transformation of Bayesian networks into possibilistic networks. They extend the definition of the consistency principle to preserve the order of interpretations and the distributions obtained after a transformation. Note that in this work, the authors focused mostly on certain existing transformations such as OT and ST while our work deal with all the transformations preserving the order of interpretations. The second work close to ours [18] addressed the commutativity of transformation with respect to some operations but its aim was to show that the obtained distributions are not identical. In our work, we are actually interested in the commutativity but only regarding the order of interpretations and events. Some of these issues were dealt with in the context of fuzzy interval analysis [9]

An interesting question is whether there exist particular probability distributions p such that the transformation operation \triangleright preserves the relative ordering between interpretations after marginalisation. A first natural idea is uniform probability distributions. Any transformation \triangleright should preserve normalisation which results in an uniform possibility distribution (where each state is associated to the possibility's degree of 1). Consequently, any event will have a possibility's degree of 1, meaning that there will not be a reversal in the order of interpretation on marginals distributions for example. Another kind of probability distributions is called "atomic bond system" [17] or big-stepped or lexicographic [6,2] probability distributions p defined by: $\forall \omega_i \in \Omega, p(\omega_i) > \sum \{p(\omega_j) : \omega_j \in \Omega \text{ and } p(\omega_j) < p(\omega_i)\}$. Clearly, if p is a big-stepped distribution then the transformation operation \triangleright preserves the ordering between interpretations after marginalisation. Note however that for both particular cases (uniform and big-stepped distribution) the ordering between non-elementary events is not preserved.

It is known that probability-possibility transformations suffer from loss of information as we move from an additive framework to a qualitative or semi-qualitative framework. But the impact on the reasoning was not yet completely studied. The results we obtained confirm that there is a loss of information at several levels regarding reasoning. But this does not mean we can do nothing with transformations. In particular, responses to MPE queries are not affected by the transformations. Which is not the case for the MAP queries unfortunately. As future works, we will study MAP inference in credal networks (based on sets of probabilities and known for their high computational complexity in comparison to Bayesian or possibilistic networks) by transforming them into possibilistic networks. This can provide good and efficient approximations for MAP inference with a better computational cost. Other open questions concern the commutativity of transformations with other definitions of conditioning and independence in the possibilistic setting.

7 Acknowledgements

This work is done with the support of a CNRS funded project PEPS FaSciDo 2015 called MAPPOS.

References

1. C. Baudrit, I. Couso, and D. Dubois. Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *International Journal of Approximate Reasoning*, 45(1):82 – 105, 2007.
2. S Benferhat, D Dubois, and H Prade. Possibilistic and standard probabilistic semantics of conditional knowledge bases. *J. Log. Comput.*, 9(6):873–895, 1999.
3. C. Borgelt, J. Gebhardt, and R. Kruse. Graphical models. In *In Proceedings of International School for the Synthesis of Expert Knowledge 98*, pages 51–68. Wiley, 2002.
4. A. Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.
5. Sébastien Destercke, Didier Dubois, and Eric Chojnacki. Transforming probability intervals into other uncertainty models. In *5th EUSFLAT Conference, Ostrava, Czech Republic, September 11-14, 2007, Volume 2: Regular Sessions*, pages 367–373, 2007.
6. D Dubois, H Fargier, and H Prade. Ordinal and probabilistic representations of acceptance. *J. Artif. Int. Res.*, 22(1):23–56, July 2004.
7. D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10(4):273–297, 2004.
8. D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.
9. D Dubois and H Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, 42(1):87–101, 1991. DP165.
10. D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In *Fuzzy Logic*, pages 103–112. 1993.
11. E. Hisdal. Conditional possibilities independence and non interaction. *Fuzzy Sets and Systems*, pages 283–297, 1978.
12. G. J. Klir and J. F. Geer. Information-preserving probability- possibility transformations: Recent developments. In *Fuzzy Logic*, pages 417–428. 1993.
13. M.-H. Masson and T. Denoeux. Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, 157(3):319–340, 2006.
14. P. Billaudel M.S Mouchaweh, M.S Bouguelid and B. Riera. Variable probability-possibility transformation. pages 417–428. September 2006.
15. H. Prade and D. Dubois. *Fuzzy Sets and Systems*, pages 258–259, 1980.
16. Y. B. Slimen, R. Ayachi, and N. B. Amor. Probability-possibility transformation: - application to bayesian and possibilistic networks. In *WILF*, volume 8256 of *Lecture Notes in Computer Science*, pages 122–130. Springer, 2013.
17. P Snow. Standard probability distributions described by rational default entailment. 1996.
18. T. Sudkamp. On probability-possibility transformations. *Fuzzy Sets and Systems*, pages 73–81, 1992.
19. L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100:9–34, 1999.