Practical Methods for Constructing Possibility Distributions

Didier Dubois, Henri Prade

Practical Methods for Constructing Possibility Distributions

Didier Dubois,* Henri Prade
IRIT, CNRS and Université de Toulouse, 31062, Toulouse Cedex 09, France

This survey paper provides an overview of existing methods for building possibility distributions. We both consider the case of qualitative possibility theory, where the scale remains ordinal, and the case of quantitative possibility theory, where the scale is the real interval [0, 1]. Methods may be order-based or similarity-based for qualitative possibility distributions, whereas statistical methods apply in the quantitative case and then possibilities encode nested random epistemic sets or upper bounds of probabilities. But distance-based approaches, or expert estimates, may be also exploited in the quantitative case.

1. INTRODUCTION

One of the key questions often raised by scientists when considering fuzzy sets is how to measure membership degrees. However, this question is hardly meaningful if no interpretive context for membership functions is provided. One such context is possibility theory, first outlined by Lotfi Zadeh in 1977.[1] Possibility distributions are the basic building blocks of possibility theory. Zadeh proposes to consider them as fuzzy set membership functions interpreted in a *disjunctive* way,[2] namely, serving as elastic constraints restricting the possible values of a *single-valued* variable. Different kinds of possibility distributions may be encountered in a variety of applications ranging from information systems and databases[3] to operations research[4] and artificial intelligence,[5] from computation with ill-known quantities represented by fuzzy intervals,[6] to the set of possible models of a possibilistic logic base[7] (see Ref. 8 for more references). Whatever the situation, having faithful elicitation or estimation methods for possibility distributions is clearly an important issue.

The idea of graded possibility was thus advocated by Zadeh in the late 1970s. But before him, the economist G. L. S. Shackle[9–11] and the philosopher David Lewis[12] did the same, albeit on the basis of concerns very different from Zadeh's. Indeed, Zadeh was mainly motivated by the representation of linguistic terms as a way of expressing uncertain and imprecise information held by humans, referring to some appropriate distance to prototypical examples; in contrast, Shackle was

*Author to whom all correspondence should be addressed; e-mail: dubois@irit.fr.

interested in modeling expectations in terms of degrees of potential surprise (which turn out to be degrees of impossibility); and Lewis advocated a comparative possibility-based view of counterfactual conditionals, where the possibility of a world depends on its similarity (or closeness) to a reference world and is represented in terms of so-called "systems of nested spheres" around this world.

Depending on the situations and the views, the concept of possibility may refer to ideas of feasibility ("it is possible *to* . . . ") or epistemic consistency ("it is possible *that* . . . "), and its evaluation is in practice either a matter of similarity (or distance)—a view recently revived by Zadeh,[13] or in terms of cost, or yet of frequency (viewing possibility as upper probability). We shall encounter these different interpretations in the following survey of techniques for constructing possibility distributions.

The paper is organized as follows: In Section 2, we first provide a refresher on possibility theory, distinguishing the qualitative and the quantitative views, and emphasizing the role of information principles in the specification of possibility distributions. Section 3 is devoted to methods for generating qualitative possibility distributions as in possibilistic logic, or when dealing with default conditionals. Section 4 provides an overview of elicitation methods for quantitative possibility distributions, based on distances, frequencies, or expert knowledge.

2. POSSIBILITY THEORY: A REFRESHER

This brief overview focuses on the possible meanings of a possibility distribution. We first review the relation between possibility distributions and fuzzy sets, before introducing possibility distributions as a representation tool for imprecise or uncertain information, together with the associated set functions for assessing the plausibility or the certainty of events. We then discuss different qualitative and quantitative scales for grading possibility, and finally address the relations between possibility and probability.

2.1. Possibility Distribution and Fuzzy Set

In his paper introducing possibility theory, Zadeh[1] starts with the representation of pieces of information of the form "X is A," where X is a parameter or attribute of interest and A is a fuzzy set on the domain of X, often representing a linguistic category (e.g., *John is Tall*, where $X = height(John)$, and A is the fuzzy set of *Tall* heights for humans). The question is then, knowing that "X is A," to determine what is the possibility distribution π_X restricting the possible values of X (also assuming we know the meaning of A, given by a $[0, 1]$-valued membership function μ_A). Then Zadeh represents the piece of information "X is A" by the elastic restriction

$$\forall u \in U, \pi_X(u) = \mu_A(u)$$

where U is the universe of discourse on which X ranges. Thus, μ_A is turned into a kind of likelihood function for X. In the above example, U is the set of human heights. Note however that π_X acts as a *disjunctive* restriction (X takes a single

value in U), while, prior to using it as above, A is a *conjunctive* fuzzy set,[2] the fuzzy set of all values more or less compatible with the meaning of A. Thus the degree of possibility that $X = u$ is evaluated as the degree of compatibility $\mu_A(u)$ of the value u with the fuzzy set A.

2.2. Representation of Imprecise Information and Specificity

In more abstract terms, π_X is a mapping from a referential U (understood as a set of mutually exclusive values for the attribute X) to a totally ordered scale L, with top denoted by 1 and bottom by 0, such as the unit interval [0, 1]. Thus any mapping from a set of elements, viewed as a mutually exclusive set of alternatives, to [0, 1] (and more generally to any totally ordered scale) can be seen as acting as an elastic restriction on the value of a single-valued variable, i.e., can be seen as a possibility distribution. Apart from the representation of ill-known numerical quantities defined on continuums, as in the human height example above, another "natural" and simple use of possibility distributions is the representation of ill-known states of affairs (or worlds, according to logicians), a concern of interest for Shackle[10] from a decision perspective.

Then U more generally stands for a (mutually exclusive) set of states of affairs (or descriptions thereof), or states, for short. The function π represents the state of knowledge of an agent (about the actual state of affairs) distinguishing what is plausible from what is less plausible, what is the normal course of things from what is not, what is surprising from what is expected. It represents a flexible restriction on what is the actual state with the following conventions:[a]

- $\pi(u) = 0$ means that state u is rejected as impossible;
- $\pi(u) = 1$ means that state u is totally possible.

If U is exhaustive, at least one of the elements of U should be the actual world, so that $\exists u, \pi(u) = 1$ (normalization). Different values may simultaneously have a degree of possibility equal to 1. In particular, extreme forms of epistemic states can be captured, namely: *complete knowledge*, where for some $u_0, \pi(u_0) = 1$ and $\pi(u) = 0, \forall u \neq u_0$ (only u_0 is possible), and *complete ignorance* where $\pi(u) = 1, \forall u \in U$ (all states are possible).

A possibility distribution π is said to be *at least as specific as* another π' if and only if for each state of affairs u, we have $\pi(u) \leq \pi'(u)$.[14] Then, π is at least as restrictive and informative as π'. This agrees with Zadeh's entailment principle that "X is A" entails "X is B," as soon as $A \subseteq B$. In the presence of pieces of knowledge coming from humans and acting as constraints, possibility theory is driven by the principle of least commitment called *minimal specificity principle*[15]. It states that any hypothesis not known to be impossible cannot be ruled out. In other words, if all we know is that "X is A," any possibility distribution

[a]The interpretation for 0 is similar to the case of probability, but Shackle's potential surprise scale is stated the other way around: 0 means possible, and the more impossible an event, the more surprising it is.

for which $\pi_X \leq \mu_A$ and $\exists u, \pi_X(u) < \mu_A(u)$ would be too restrictive, since we have no further information that could support the latter strict inequality. Hence, $\pi_X = \mu_A$ is the right representation, if we have no further information. The minimal specificity principle justifies the use of the *minimum*-based combination principle of n pieces of information of the form "X is A_i," in approximate reasoning,[16] since $\pi_X = \min_{i=1}^{n} \mu_{A_i}$ is the largest possibility distribution such that we have $\pi \leq \mu_{A_i}, \forall i = 1, \ldots, n$.

Sometimes, the opposite principle must be used. This is when we possess statistical information that represents data and not knowledge. In this case, we consider the most specific possibility distribution enclosing the data, assuming, like in probability density estimation, that what has not been observed is impossible.[17] This is similar to the closed-world assumption.

2.3. Possibilistic Set Functions

Given a simple query of the form "does event A occur?" where A is a subset of states, the response to the query can be obtained by computing degrees of possibility and necessity, respectively (assuming the possibility scale $L = [0, 1]$):

$$\Pi(A) = \sup_{u \in A} \pi(u); N(A) = \inf_{u \notin A} (1 - \pi(u)).$$

$\Pi(A)$ evaluates to what extent A is logically consistent with π, whereas $N(A)$ evaluates to what extent A is certainly implied by π. The possibility-necessity duality says that a proposition is certain if its opposite is impossible, and this is expressed by

$$N(A) = 1 - \Pi(A^c),$$

where A^c is the complement of A. Generally, $\Pi(U) = N(U) = 1$ and $\Pi(\emptyset) = N(\emptyset) = 0$. Possibility measures satisfy the basic "maxitivity" property

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)).$$

Necessity measures satisfy a "minitivity axiom" dual to that of possibility measures, namely

$$N(A \cap B) = \min(N(A), N(B)),$$

expressing that being certain that $A \cap B$ is the same as being certain of A and of B.

Human knowledge is often expressed in a declarative way, using statements to which belief degrees are attached. This format corresponds to expressing constraints with which the world is supposed to comply. Certainty-qualified pieces of uncertain information of the form "(X is A) is certain to degree α" can then be modeled by the constraint $N(A) \geq \alpha$. The least specific possibility distribution reflecting this

information is[15]:

$$\pi_{(A,\alpha)}(u) = \begin{cases} 1, \text{ if } u \in A \\ 1 - \alpha \text{ otherwise.} \end{cases} \tag{1}$$

Acquiring further pieces of knowledge consistent with the former leads to updating $\pi_{(A,\alpha)}$ into some $\pi < \pi_{(A,\alpha)}$. Another example where the principle of minimal specificity is useful is when defining the notion of conditioning in possibility theory. The most usual form respects an equation of the form

$$\Pi(A \cap B) = \Pi(A|B) \star \Pi(B), \quad N(A|B) : 1 - \Pi(A^c|B), \tag{2}$$

where $\star$ is a t-norm and $B \neq \emptyset$. The most justified choices of $\star$ are min and product.[18] In the case of product, it looks like probabilistic conditioning applied to possibility measures and corresponds to Dempster conditioning.[19] Using min, the above definition (3) does not yield a unique conditional possibility. Then the idea is to use the least specific possibility measure respecting (2), i.e.,

$$\Pi(A|B) = \begin{cases} \Pi(A \cap B) \text{ if } \Pi(A \cap B) < \Pi(B), \\ 1 \text{ otherwise.} \end{cases} \tag{3}$$

Apart from Π and N, a measure of *guaranteed possibility* or *sufficiency* can be defined[20,21]: $\Delta(A) = \inf_{u \in A} \pi(u)$. It estimates to what extent *all* states in A are actually possible according to evidence. $\Delta(A)$ can be used as a degree of evidential support for A. In contrast, Π appears to be a measure of *potential* possibility. Uncertain statements of the form "B is possible to degree β" often mean that all realizations of B are possible to degree β. They can then be modeled by the constraint $\Delta(B) \geq \beta$. It corresponds to the idea of observed evidence. This type of information is better exploited by an informational principle opposite to the one discussed above (minimal specificity would give nothing). The most specific distribution $\delta_{(B,\beta)}$ in agreement with $\Delta(B) \geq \beta$ is

$$\delta_{(B,\beta)}(u) = \left\{ \begin{array}{ll} \beta, & \text{if } u \in B \\ 0 & \text{otherwise.} \end{array} \right\}$$

Acquiring further pieces of evidence leads to updating $\delta_{(B,\beta)}$ into some wider distribution $\delta > \delta_{(B,\beta)}$.[21]

2.4. Different Scales for Graded Possibility

There are several representations of epistemic states that are in agreement with the above setting such as well-ordered partitions,[22] Lewis' systems of spheres,[12,23] Spohn's "ordinal conditional functions" (OCF)[22,24] (also called ranking functions[25]), and possibilities viewed as upper probabilities. But all these representations of epistemic states do not have the same expressive power. They range from purely qualitative to quantitative possibility distributions, using weak orders, qualitative

scales, integers, and reals. In fact, we can distinguish several representation settings according to the expressiveness of the scale used[26]:

1. The purely ordinal setting, where an epistemic state on a set of possible worlds is simply encoded by means of a total preorder $\succeq$, telling which worlds are more normal, less surprising than other ones. The quotient set $U/\sim$, built from the equivalence relation $\sim$ extracted from $\succeq$, forms a *well-ordered partition* $E_1, \ldots, E_k$ such that the greater the index i, the less plausible or the less likely the possible states in E_i. In that case the comparative possibility relation $\succeq_\Pi$ is such that $A \succeq_\Pi B$ if and only if $\exists u_1 \in A, \forall u_2 \in B, u_1 \succeq u_2$. This is the setting used by Lewis[12] and by Grove[23] and Gärdenfors[27] when modeling belief revision. Only possibility measures can account for such relations.[28]
2. The qualitative finite setting, with possibility degrees in a finite totally ordered scale: $L = \{\alpha_0 = 1 > \alpha_1 > \ldots > \alpha_{m-1} > 0\}$. This setting has a classificatory flavor, as we assign each event to a class in a finite totally ordered set thereof, corresponding to the finite scale of possibility levels. It is used in possibilistic logic.[7] However, note that the previous purely ordinal representation is less expressive than the qualitative encoding of a possibility distribution on a totally ordered scale, as the former cannot express absolute impossibility.
3. The denumerable setting, using a scale of *powers* $L = \{\alpha^0 = 1 > \alpha^1 > \ldots > \alpha^i > \ldots, 0\}$, for some $\alpha \in (0, 1)$. This is isomorphic to the use of integers in ranking functions by Spohn,[25] where the set of natural integers is used as a disbelief scale.
4. The dense ordinal scale setting using $L = [0, 1]$, seen as an ordinal scale. In this case, the possibility distribution Π is defined up to any monotone increasing transformation $f : [0, 1] \to [0, 1]$, $f(0) = 0$, $f(1) = 1$. This setting is also used in possibilistic logic.[7]
5. The dense absolute setting, where $L = [0, 1]$, seen as a genuine numerical scale equipped with product. In this case, a possibility measure can be viewed as a special case of Shafer's plausibility function,[29] actually a consonant one, and $1 - \pi$ as a potential surprise function in the sense of Shackle.[11]

2.5. Quantitative Possibilities and Their Links with Probabilities

The idea of a link between graded possibility and probability is natural since both acts as modalities for expressing some form of uncertainty. This link may be stated under the form of a consistency principle[1] stating that "what is possible may not be probable and what is improbable need not be impossible." Proceeding further, we may consider that what is probable should be possible, and what is necessarily (certainly) the case should be probable as well. This amounts to writing $N \leq P \leq \Pi$, where N, P, and Π are, respectively, a necessity, a probability, and a possibility measure (Ref. 30, p. 138).

Let π be a possibility distribution where $\pi(u) \in [0, 1]$. Let $\mathbf{P}(\pi)$ be the never empty set of probability measures P such that $P \leq \Pi$, i.e. $\forall A \subseteq U$, $P(A) \leq \Pi(A)$ (equivalently, $P \geq N$). Then the possibility measure Π coincides with the upper probability function P^* such that $P^*(A) = \sup\{P(A), P \in \mathbf{P}(\pi)\}$, whereas the necessity measure N is the lower probability function P_* such that $P_*(A) = \inf\{P(A), P \in \mathbf{P}(\pi)\}$; see Refs. 31, 32 for details. P and π are said to be compatible if $P \in \mathbf{P}(\pi)$. So, Π and N are coherent upper and lower probabilities in the sense of Walley,[33] as already pointed out very early by Giles.[34] The connection between possibility measures and imprecise probabilistic reasoning is especially interesting for the efficient representation of nonparametric families of probability functions, and it makes sense even in the scope of modeling linguistic information.[35]

A possibility measure can thus be computed from a set of nested confidence subsets $\{A_1, A_2, \ldots, A_k\}$, where $A_i \subset A_{i+1}, i = 1 \ldots, k-1$. To each confidence subset A_i is attached a positive confidence level λ_i interpreted as a lower bound of $P(A_i)$, hence a necessity degree. The pair (A_i, λ_i) can be viewed as a certainty-qualified statement that generates a possibility distribution π_i, as recalled above. The corresponding possibility distribution is obtained by intersecting fuzzy sets like those in Equation 1:

$$\pi(u) = \min_{i=1,\ldots,k} \pi_i(u) = \begin{cases} 1 \text{ if } u \in A_1 \\ 1 - \lambda_{j-1} \text{ if } j = \max\{i : u \notin A_i\} > 1. \end{cases} \tag{4}$$

The information modeled by π can also be viewed as a nested random set

$$\{(A_i, m(A_i)), i = 1, \ldots, k\},$$

associated with a belief function,[36] letting $m(A_i) = \lambda_i - \lambda_{i-1}$.[37] This framework allows for imprecision (reflected by the size of the A_is) and uncertainty (the $m(A_i)$s). And $m(A_i)$ is the probability that the agent only knows that A_i contains the actual state (it is not $P(A_i)$). The random set view of possibility theory is well adapted to the idea of imprecise statistical data, as developed in Section 4. Conversely, if a belief function is consonant then its contour function $\pi(u) = \sum_{i:u\in A_i} m(A_i)$ is sufficient to recover the belief function, where m is its basic probability assignment ($\sum_i m(A_i) = 1$), and the A_is are both the nested focal elements associated with m, and the level cuts of π.

REMARK *1. Let us mention another possible kind of link between very small probabilities and possibilities. This interpretation has been pointed out by Spohn*[24] *for his integer-valued ranking functions* κ *ranging from 0 to* $+\infty$ *(0 meaning full possibility, and* $+\infty$ *full impossibility), where* $\kappa(A)$ *may be thought of as a degree of disbelief modeled by a kind of cost. Namely* $\kappa(A) = k$ *is interpreted as a small probability of the form* ϵ^k *with* $\epsilon \ll 1$ *(e.g.,* $P(A) = 10^{-7}$*, when* $\epsilon = 0.1$*, and* $k = 7$*), i.e., the probability of a rare event. Indeed if A has a small probability with the order of magnitude* ϵ^k*, and B is another event with a small probability with the order of magnitude* ϵ^n*, the order of magnitude of the probability* $P(A \cup B)$ *is* $\epsilon^{\min(k,n)}$*, which mirrors the maxitivity decomposition property of possibility measures, up to a rescaling from* $[0, +\infty)$ *to* $[0, 1]$[38]*. It suggests an interpretation of possibility (and necessity) measures in terms of probabilities of rare events.*

3. CONSTRUCTION METHODS FOR QUALITATIVE POSSIBILITY DISTRIBUTIONS

The elicitation of qualitative possibility distributions is made easier by the qualitative nature of possibility degrees. Indeed, even in a dense ordinal scale $L = [0, 1]$, the precise values of the degrees do not matter, only their relative values are important as expressing strict inequalities between possibility levels. In fact, it basically amounts to determining a well-ordered partition.

In a purely ordinal setting, a possibility ordering is a complete preorder of states denoted by $\geq_\pi$, which determines a well-ordered partition $\{E_1, \ldots, E_k\}$ of U. It is the comparative counterpart of a possibility distribution π, i.e., $u \geq_\pi u'$ if and only if $\pi(u) \geq \pi(u')$. By convention E_1 contains the most plausible (or normal), or the most satisfactory (or acceptable) states, E_k the least plausible (or most surprising), or the least satisfactory ones, depending if we are modeling knowledge, or preferences. Ordinal counterparts of possibility and necessity measures[28] are defined as follows: $\{u\} \geq_\Pi \emptyset$ for all $u \in U$ and

$$A \geq_\Pi B \text{ if and only if } \max(A) \geq_\pi \max(B)$$
$$A \geq_N B \text{ if and only if } \max(B^c) \geq_\pi \max(A^c).$$

Possibility relations $\geq_\Pi$ are those of Lewis.[12] They satisfy the characteristic property

$$A \geq_\Pi B \text{ implies } C \cup A \geq_\Pi C \cup B,$$

while necessity relations can also be defined as $A \geq_N B$ if and only if $B^c \geq_\Pi A^c$ and satisfy a similar property:

$$A \geq_N B \text{ implies } C \cap A \geq_N C \cap B.$$

Necessity relations coincide with epistemic entrenchment relations in the sense of belief revision theory.[27,38] In particular, the assertion $A >_\Pi A^c$ expresses the acceptance of A[39] and is the qualitative counterpart of $N(A) > 0$. This qualitative setting enables qualitative possibility distributions to be derived either from a set of certainty-qualified propositions, or from a set of conditional statements.

3.1. Certainty-Qualified Propositions

When an agent states beliefs with their (relative) strengths, it is more natural to expect that ordinal information, rather than truly numerical information, is supplied. This gives birth to a knowledge base in the sense of possibilistic logic,[7] i.e., a set of weighted statements $K = \{(A_i, \alpha_i) : i = 1, \ldots, m\}$, each of them representing a constraint $N(A_i) \geq \alpha_i$, where A_i represents a subset of possible states or interpretations and α_i is the associated certainty level (or priority level) belonging to a denumerable ordinal scale. Such a base K is semantically associated with the possibility distribution in (4), where we no longer assume nested events:

$$\pi_K(u) = \min_{i=1,\ldots,m} \pi_{(A_i,\alpha_i)}(u) = \min_{i=1,\ldots,m} \max(\mu_{A_i}(u), 1 - \alpha_i)$$

and μ_{A_i} is the characteristic function of the subset A_i. Besides, the α_is may also have a similarity flavor when some pair (A_i, α_i) correspond to the level-cuts of fuzzy subsets.[40,41]

Let us mention that a similar construction can be made in an additive setting where each formula is associated with a cost (in $\mathbb{N} \cup \{+\infty\}$), the weight (cost)

attached to an interpretation being the sum of the costs of the formulas in the base violated by the interpretation, as in penalty logic.[42] The so-called "cost of consistency" of a formula is then defined as the minimum of the weights of its models. It is nothing but a ranking function (OCF) in the sense of Spohn,[24] the counterpart of a possibility measure defined on $\mathbb{N} \cup \{+\infty\}$, where now 0 expresses full possibility (free violation), and $+\infty$ complete impossibility (a price that cannot be paid). However, this view gives a more quantitative flavor to the construction, thus moving from a qualitative setting to a numerical one.

The construction of π_K from the collection of statements in K clearly relies on the application of the minimal specificity principle. As mentioned in the previous section, a dual principle may be more appropriate when we start from data, rather than constraints excluding impossible states. Assume that we have a collection of weighted data $D = \{(B_j, \beta_j),\ j = 1, \ldots, n\}$, understood as $\Delta(B_j) \geq \beta_j$, where the β_js belong to an ordinal scale and reflect, e.g., some similarity-based relevance of the data. Then by virtue of maximal specificity, we get the lower possibility distribution (which needs not to be normalized):

$$\delta_D(u) = \max_{j=1,\ldots,n} \delta_{(B_j,\beta_j)}(u) = \max_{j=1,\ldots,n} \min(\mu_{B_i}(u), \beta_j).$$

Note that this expression takes the form of the kind of fuzzy conclusions (prior to defuzzification) obtained from Mamdani fuzzy rule-based systems.[43]

3.2. Indicative Conditionals

Besides, there exists yet another method to obtain a qualitative possibility distribution, starting from a set of *conditionals*, rather than from a set of lower bounds on the necessity, or the guaranteed possibility, of a collection of subsets. This method was originally invented for stratifying a set of default rules to design proper methods for handling exception-tolerant reasoning about incompletely described cases; see, e.g., Ref. 44. A default rule "if A then B, generally," denoted $A \rightsquigarrow B$, is then understood formally as the conditional constraint

$$\Pi(A \cap B) > \Pi(A \cap B^c)$$

on a possibility measure Π, expressing that the examples of the rule (the situations where A and B hold) are more plausible than its counterexamples (the situations where A holds and B does not). It is equivalent to the conditional statement $N(B|A) > 0$. Remember that, in contrast, the probabilistic interpretation is such that $P(A \cap B) > P(A \cap B^c)$ if and only if $P(B|A) > 1/2$.

The above possibilistic constraint can be equivalently expressed in terms of a mere comparative possibility relation, namely $A \cap B >_\Pi A \cap B^c$. Any finite *consistent* set of constraints of the form $A_k \cap B_k >_\Pi A_k \cap B_k^c$, representing a set of defaults $\Delta = \{A_k \rightsquigarrow B_k, k = 1, \ldots, r\}$, is compatible with a nonempty family of relations $>_\Pi$ and determines a partially defined ranking $>_\pi$ on U that can be completed according to the principle of minimal specificity. This principle assigns to

each state u the highest possibility level (in forming a well-ordered partition of U) without violating the constraints. It defines a unique complete preorder.[44] Let $E_1, \ldots, E_k$ be the obtained partition. Then $u >_\pi u'$ if $u \in E_i$ and $u' \in E_j$ with $i < j$, whereas $u \sim_\pi u'$ if $u \in E_i$ and $u' \in E_i$ (where $\sim_\pi$ means $\geq_\pi$ and $\leq_\pi$).

A numerical counterpart to $>_\pi$ on a denumerable finite scale can be defined by $\pi(u) = \frac{k+1-j}{k}$ if $u \in E_j$, $j = 1, \ldots, k$.[44] Note that it is purely a matter of convenience to use a numerical scale, and any other numerical counterpart such that $\pi(u) > \pi(u')$ iff $u >_\pi u'$ will work as well. Namely, the range of π is used as an ordinal scale. This approach has an infinitesimal probability counterpart, namely, a procedure called system Z.[45] It has been refined by the numerical system Z^+,[46] whose possibilistic counterpart corresponds to the handling of "strengthened" constraints of the form $\Pi(A_j \cap B_j) > \rho_j \cdot \Pi(A_j \cap B_j^c)$, where $\rho_j \geq 1$. This approach can also be expressed in terms of conditioning in the setting of Spohn's ranking functions. Note that the latter methods were intended to stratify default knowledge bases rather than to explicitly derive possibility distributions.

4. CONSTRUCTION METHODS FOR QUANTITATIVE POSSIBILITY DISTRIBUTIONS

The construction of possibility distributions in the quantitative setting either rely on numerical similarity or exploit the connection between probability and possibility inspired by Zadeh[1] according to whom what is probable must be possible, which is understood here by the inequality $\Pi(A) \geq P(A)$, for all measurable subsets A. In the first case, possibility is viewed as a form of renormalized distance to most plausible values. In the second case, it means that we can derive possibility distributions from statistical data or from subjective probability elicitation methods.

4.1. Possibility as Similarity

In his approach to the non-Boolean representation of natural language categories, Zadeh[2] uses membership functions representing the extensions of fuzzy predicates to derive possibility distributions, as recalled in Section 2.1. If we know the membership function μ_{Tall} of *Tall* on the scale of human heights, then the piece of information *John is Tall*, accepted as being true, can be represented by a possibility distribution $\pi_{hgt(John)}$ equated with μ_{Tall}:

$$\pi_{hgt(John)}(h) = \mu_{Tall}(h).$$

In other words, the measurement of possibility degrees comes down to the measurement of membership functions of linguistic terms. However, in such a situation, $\mu_{Tall}(h)$ is often constructed as a function of the distance between the value a and the closest height $\hat{h}$ that can be considered prototypical for *Tall*, i.e., $\mu_{Tall}(\hat{h}) = 1$, for instance,

$$\mu_{Tall}(h) = f(d(h, \hat{h})) \tag{5}$$

where f is a nonnegative, decreasing function such that $f(0) = 1$, for instance $f(u) = \frac{1}{1+u}$, and $d(h, \hat{h}) = \min\{d(h, x) : \mu_{Tall}(x) = 1\}$, where d is a distance. Sudkamp[47] points out that conversely, given a possibility distribution π, the two-place function $\delta(x, y) = |\pi(x) - \pi(y)|$ is a pseudodistance indeed.

Results of fuzzy clustering methods can be interpreted as distance-based membership functions. Alternatively, one may define a fuzzy set F from a crisp set A of prototypes of μ_{Tall} and a similarity relation $S(x, y)$ on the height scale, such that $S(x, x) = 1$ (then $1 - S(x, y)$ is akin to a distance). Ruspini[48] proposes to define the membership function as a kind of upper approximation of A:

$$\mu_F(h) = \max_{u \in A} S(u, h).$$

Then A stands as the core of the fuzzy set F. We refer the reader to the survey by Türksen and Bilgic[49] for membership degree elicitation using measurement methods outside the possibility theory view, and more recently to papers by Marchant[50,51].

Besides, the idea of relating plausibility and distance also pervades the probabilistic literature: The use of normal distributions as likelihood functions can be viewed as a way to define degrees of likelihood via the Euclidean distance between a given number and the most likely value (which in that case coincides with the mean value of the distribution). In the neurofuzzy literature, one often uses Gaussian membership functions of the form (5) with $f = e^{-x^2}$.

4.2. Statistical Interpretations of Possibility Distributions

The use of possibility distributions seems to range far beyond the linguistic point of view advocated by Zadeh.[2] Namely, the use of (normalized) membership functions interpreted as ruling out the more or less impossible values of an ill-known quantity X, as well as the maxitivity axiom of possibility measures, are actually often found in the statistical literature, in connection with the non-Kolmogorovian aspects of statistics, namely the maximum likelihood principle, the comparison of probability distributions in terms of dispersion, and the notion of confidence interval; see Refs. 52, 53, 54 for surveys of such connections between probability and possibility. In this section, we focus on the derivation of possibility distributions from a (finite) set of statistical data.

4.2.1. Interval Data

It is useful to cast the problem in a more general setting, namely the one of set-valued data, and the theory of random sets.[55–57] Consider a random variable X and a (multi)set of data reporting the results of some experiments under the form of intervals $\mathcal{D} = \{I_i : i = 1, \ldots, n\}$ subsets of a real interval $U = [a, b]$. In general, due to randomness, one cannot expect this set of intervals to be nested. Representing it by a possibility distribution will result in an approximation to this information. Strictly speaking, what is needed to represent this data set exactly is a random set

defined by a mass function $m : 2^{[a,b]} \to [0, 1]$ such that

$$m(E) = \frac{|\{I_i : E = I_i\}|}{n}, \forall E \subseteq [a, b] \tag{6}$$

Note that this expression is formally related to a belief function $Bel(A) = \sum_{E \subseteq A} m(E)$ of Shafer.[36] In particular, each focal set E with $m(E) > 0$ represents incomplete information, namely that some $x_i \in I_i$ should have been observed as the result of the ith experiment, but only an imprecise representation of this observation could be obtained in the form of I_i. However, in the theory of evidence, Shafer assumes that $m(E)$ is a subjective probability (the probability that the set E is a faithful representation of an agent's knowledge about X). The interval data are more in conformity with Dempster[19] view, since $m(E)$ is the frequency of observing E.

In fact, $\mathcal{D} = \{I_i : i = 1, \ldots, n\}$ is interpreted as an *epistemic random set,*[55] i.e., it describes an ill-known standard random variable. It represents the (finite, hence nonconvex) set of probabilities obtained by all selections of values in the intervals of $\mathcal{D}$. Let $d^k = \{x_1^k, \ldots, x_n^k\}$ represent a precise data set compatible with $\mathcal{D}$ in the sense that $x_i^k \in I_i, i = 1, \ldots, n$. This is denoted by $d^k \in \mathcal{D}$. Moreover, the belief function $Bel(A)$ is a lower frequency of A, whereas the plausibility degree $Pl(A) = \sum_{E \cap A \neq \emptyset} m(E)$ is an upper frequency. Let $f^k(a)$ be the frequency of $u = x_i^k$ in d^k. Then

$$Bel(A) = \min_{d^k \in \mathcal{D}} \sum_{u \in A} f^k(u); \quad Pl(A) = \max_{d^k \in \mathcal{D}} \sum_{u \in A} f^k(u).$$

(See Refs. 58, 59, 56 for more on statistics with interval data.)

A straightforward way of deriving a possibility distribution from such statistical data is to consider what Shafer[36] called the *contour function* of m (actually, the one-point coverage function of the random set):

$$\pi_*(a) = \sum_{a \in E} m(E).$$

Note that this is only a partial view of the data, as it is generally not possible to reconstruct m from π_*. This view of possibility distributions and fuzzy sets as random sets was very early pointed out by Kampé de Feriet[60] and Goodman.[61] From a possibility theory point of view, it has some drawbacks:

- π_* is generally not normalized, hence not a proper possibility distribution (unless the data are not conflicting : $\bigcap_{i=1}^n I_i \neq \emptyset$). For instance, $\pi_* = m$ is a probability distribution when data are precise.
- Even when it is normalized, the interval $[N_*(A), \Pi_*(A)]$ determined by π_* is the widest interval of this form contained in $[Bel(A), Pl(A)]$[62].

One may be more interested to get the narrowest ranges $[N(A), \Pi(A)]$ *containing* intervals $[Bel(A), Pl(A)]$, as being safer; see Ref. 62 for an extensive discussion of this difficult problem whose solution is not unique. The idea, first

suggested in Ref. 63 is to choose a family $\mathcal{F} = \{E_1 \subseteq \ldots \subseteq E_q\}$ of nested intervals such that $I_i \subseteq E_q$ for all intervals I_i, and $I_i \subseteq E_1$ for at least one I_i. Then it is easy to compute a nested random set $m_{\mathcal{F}}$, as follows: for each interval I_i let $\alpha(i) = \min\{j : I_i \subseteq E_j\}$, such that $E_{\alpha(i)}$ is the most narrow interval in $\mathcal{F}$ containing I_i. Then let $m_{\mathcal{F}}(E_j) = \sum_{E:E=I_i, j=\alpha(i)} m(E)$, where m is the original mass function given by (6). An upper possibility distribution $\pi_{\mathcal{F}}$ is derived such that

$$\pi_{\mathcal{F}}(a) = \sum_{a \in E_i} m_{\mathcal{F}}(E_j)$$

in the sense that $[Bel(A), Pl(A)] \subseteq [N_{\mathcal{F}}(A), \Pi_{\mathcal{F}}(A)]$. The difficult point is to choose a proper family of nested set $\mathcal{F}$. Clearly, the intervals in $\mathcal{F}$ should be as narrow as possible. One may, for instance, choose $\mathcal{F}$ in the family of cuts of π_*.

Interestingly, the random set $\{(E_j, m_{\mathcal{F}}(E_j)) : j = 1, \ldots, q\}$ can be viewed as a nested histogram, which is what is expected with empirical possibility distributions (while building a standard histogram comes down to choosing a partition of $[a, b]$).

4.2.2. From Large Precise Data Sets to Possibility Distributions

If we consider the special case of a standard point-valued data set, there does not exist a lower possibility distribution, but it is possible to derive an upper possibility distribution using a nested histogram. Of course, we lose much information, as we replace precise values by sets containing them. However, the problem of finding an optimal upper distribution has a solution known for a long time.[37,64] Consider a histogram $\mathcal{H}$ made of a partition $\{H_1, \ldots, H_n\}$ of $[a, b]$ with corresponding probabilities $p_1 > p_2 > \ldots > p_n$. Note that it is, strictly speaking, a special case of random set with disjoint realizations. Then, there is a most specific possibility distribution π^* dominating the probability distribution, called *optimal transformation*, namely

$$\forall a \in H_i, \pi^*(a) = \sum_{j \geq i} p_j. \tag{7}$$

Indeed, one can check that $P(A) \in [N^*(A), \Pi^*(A)]$ and $\Pi^*(\bigcup_{i=1}^{j} H_j) = P(\bigcup_{i=1}^{j} H_j)$. The distribution π^* is known as the *Lorentz curve* of the vector $(p_1, p_2, \ldots, p_n)$. In fact, the main reason why this transformation is interesting is that it provides a systematic method for comparing probability distributions in terms of their relative peakedness (or dispersion). Namely, it has been shown that if π_p^* and π_q^* are optimal transformations of distributions p and q (sharing the same order of elements), and $\pi_p^* < \pi_q^*$ (the former is more informative than the latter), then $-\sum_{i=1}^{n} p_i \ln p_i < -\sum_{i=1}^{n} q_i \ln q_i$, and this property holds for all entropies.[65]

Note that many authors suggest another transformation consisting in a mere renormalization of the probability distribution in the style of possibility theory,

namely

$$\pi^r(a) = \frac{p_i}{p_1}, \text{ if } a \in H_i. \tag{8}$$

However, it was already indicated in Ref. 30 (p. 259) that the inequality $\Pi^r(A) \geq P(A)$ may fail to hold for some events A. In fact, for $n = 3$, one can prove the following:

PROPOSITION 1. *Consider a probability distribution $p_1 \geq p_2 \geq p_3$ on a three-element set $\{1, 2, 3\}$. Then $\Pi^r(A) < P(A)$ for some A if and only if $p_1 > 0.5$ and $p_2 < p_1(1 - p_1)$.*

Proof. The only problematic event is $\{2, 3\}$ as $\Pi^r(A) \geq P(A)$ obviously for other events. Noticing that $p_1 = 1 - p_2 - p_3$, the condition $\Pi^r(\{2, 3\}) = \frac{p_2}{p_1} < P(\{2, 3\})$ boils down to the inequality $p_2 < p_1(1 - p_1)$. Moreover, the condition $p_2 \geq p_3$ is actually $p_2 \geq 1 - p_1 - p_2$, i.e., $p_2 \geq \frac{1-p_1}{2}$. So we need $\frac{1-p_1}{2} < p_1(1 - p_1)$, i.e., $p_1 > 0.5$. □

For instance, take $p_1 = 0.6$, $p_2 = p_3 = 0.2$; then $\Pi^r(\{2, 3\}) = 1/3 < P(\{2, 3\}) = 0.4$. In the case of more than three elements, one may find probability values $p_1 \geq \ldots \geq p_n$, such that $\frac{p_i}{p_1} < P(\{i, \ldots, n\})$, for all $i = 2, \ldots, n - 1$. It is sufficient to have $p_1 > 0.5$ and then to choose $0 < p_i < p_1(1 - \sum_{j=1}^{i-1} p_j)$, $i = 2, \ldots, n - 1$ in this order, making sure that $p_n \leq p_{n-1}$.

4.2.3. Scarce Precise Data

Another case when a possibilistic representation can be envisaged is when the data set $\mathcal{D} = \{x_i : i = 1, \ldots, n\}$ is too small. Applying estimation methods to compute the probability distribution leads to large confidence intervals. Namely, if $p(x|\theta)$ is the density to be estimated via a parameter θ, then we get confidence intervals J_β for θ with confidence level $\beta \in [0, 1]$. Usually, $\beta = 0.95$ is selected. The interval J_β is random and contains θ with probability at least β. As the confidence intervals are nested, this family of confidence intervals can be modeled by a possibility distribution over the values of θ, which comes down to a possibility distribution over probabilistic models $p(x|\theta)$. This result is similar to the one we get from fuzzy probability qualification of a linguistic statement of the form "*X is F is* $\tilde{p}$" where $\tilde{p}$ is a fuzzy interval on the probability scale. According to Zadeh,[2] this piece of information comes down to computing the possibility distribution π over probability measures P (on the range of X) for which $\pi(P) = \mu_{\tilde{p}}(P(F))$ where $P(F)$ is the scalar probability of the fuzzy event F.

4.2.3.1. Finite setting. In the case of a multinomial setting with n states, the identification of the probabilities p_i of states i based on observation frequencies f_i also yields confidence intervals. Fixing the confidence level, one gets probability intervals $[l_i, u_i]$ likely to contain the true probabilities p_i. Such probability intervals lead to upper (and lower) probabilities of events that are submodular (and

supermodular), a property far weaker than the property of possibility and necessity measures.[66] They can be approximated by possibility and necessity measures as done by de Campos and Huete[67], Masson and Denoeux[68]; see also Destercke et al.[69].

De Campos and Huete consider a finite set of n possibilities, and a small sample of N observations, where N_i is the number of observations of class i. Maximum likelihood gives probabilities $p_i = \frac{N_i}{N}$, and the statistical literature enables bounds $l_i \leq p_i \leq u_i$ to be computed as $p_i \pm c_\epsilon \sqrt{\frac{p_i(1-p_i)}{N}}$ (if inside $[0, 1]$), where c_ϵ is the appropriate percentile of the standard normal distribution. These bounds have the peculiarity that the rankings of the lower bounds, of the upper bounds, and of the p_is are the same. Based on this ranking, the authors consider extending possibility-probability transformations (7) and (8) to probability intervals (as well as the converse of the pignistic transform (11) presented later in this paper) in such a way as to verify a number of expected properties:

1. The obtained possibility degrees for each class should be in agreement with the ranking provided by the sample sizes N_i;
2. The wider the intervals $[l_i, u_i]$, the less specific the possibility distribution;
3. The larger the sample size N, the more specific the possibility distribution;
4. The possibility distribution obtained from any probability assignment in the intervals and in agreement with the sample size should be more specific than the possibility distribution obtained from the intervals.

These transformations are simple to compute. In contrast, Masson and Denoeux[68] consider the probability intervals as being partially ordered and consider the transforms of all probability distributions consistent with these intervals according to all rankings extending the partial order. The obtained possibility distribution is covering all of them. This method is combinatorially more demanding.

4.2.3.2. Continuous setting. An extreme case of scarce data is when a single observation $x = x_0$ on the real line has been obtained. Mauris[70] has shown that if we assume that the generation process is based on a unimodal distribution with mode $M = x_0$, it is possible to compute a possibility distribution whose associated necessity functions bounds the probability of events from below. This perhaps surprising fact comes from the following result[71] used by Mauris: For any value $t > 1$ and any interval $I_t = [x - |x|t, x + |x|t]$ containing the mode M of the distribution, it holds that $P(I_t) \geq 1 - \frac{2}{1-t}, \forall t > 1$. Then if the observed value $x_0 > 0$ is supposed to coincide with the mode of the distribution, we can derive a possibility distribution

$$\pi(x_0(1-t)) = \pi(x_0(1+t)) = \begin{cases} \frac{2}{1-t} \text{ if } t > 1 \\ 1 \text{ otherwise.} \end{cases}$$

This is done by interpreting $1 - \frac{2}{1-t}$ as a degree of necessity and by applying the minimal specificity principle to all such inequality constraints. Then, we know that whatever the underlying probability measure with mode x_0, we get $P(A) \geq N(A)$, where N is constructed from π. The above result of Mauris[70] can be improved if more assumptions are made (symmetry, shape of the distribution) or if several observations

obtained. Also, if the variable of interest is known to be bounded, i.e., to lie inside an interval $[a, b]$, Dubois et al.[73] have shown that the triangular possibility distribution with mode x_0 and support $[a, b]$ also dominates the probability of any event A for all unimodal probability distributions with mode x_0 and support in $[a, b]$ (including uniform ones); see Mauris[53,54] for a more extensive view of the role of possibility distributions in statistics (evaluation of dispersion, estimation methods, etc.).

4.2.4. Possibility Measures and Cumulative Distributions

Possibility distributions, when related to probability measures, are closely related to cumulative distributions, as already suggested by expression (7). Namely, given a family $I_t = [a_t, b_t], t \in [0, 1]$ of nested intervals, such that $t < s$ implies $I_s \subset I_t, I_1 = \{\hat{x}\}$, and a probability measure P whose support lies in $[a_0, b_0]$, letting

$$\pi(a_t) = \pi(b_t) = 1 - P(I_t), t \in [0, 1]$$

yields a possibility distribution (it is the membership function of a fuzzy interval) that is compatible with P. Now, $1 - P(I_t) = P((-\infty, a_t)) + P((b_t, +\infty))$ making it clear that the possibility distribution coincides with a two-sided cumulative distribution function. Choosing $I_t = \{x : p(x) \geq t\}$ for $t \in [0, \sup p]$, where p is the density of P, one gets the most specific possibility distribution compatible with P[72]. It has the same shape as p and $\hat{x}$ is the mode of p. It is the continuous counterpart of equation (7). It provides a faithfull description of the dispersion of P[72].

Conversely, given a possibility distribution in π the form of a fuzzy interval, then the set of probability measures $\mathcal{P}(\pi)$ dominated by its possibility measure Π is equal to $\{P : P(\pi_\alpha) \geq 1 - \alpha, \forall \alpha \in (0, 1]\}$, where $\pi_\alpha = \{x : \pi(x) \geq \alpha\}$, the α-cut of π, is a closed interval $[a_\alpha, b_\alpha]$.[73,74]

When π is an increasing function, it is generally the cumulative distribution of a unique probability measure such that $P((-\infty, x)) = \pi(x)$. Otherwise, a possibility distribution π does not determine a unique probability distribution P, contrary to the situation with usual continuous cumulative distributions. Namely, there is not a unique probability measure such that $\alpha = 1 - P(\pi_\alpha), \forall \alpha \in (0, 1]$. To show there are many probability measures such that $\alpha = 1 - P(\pi_\alpha)$, first consider the upper and lower distributions functions F^+ and F^- determined by π as follows:

$$F^+(x) = \Pi((-\infty, x]), \quad F^-(x) = N((-\infty, x]) \tag{9}$$

It should be clear that if P^+ and P^- are the probability measures associated with cumulative distributions F^+ and F^-, we do have that $\alpha = 1 - P^+(\pi_\alpha)$, and $\alpha = 1 - P^-(\pi_\alpha), \forall \alpha \in (0, 1]$. Indeed, $1 - P^+(\pi_\alpha) = P^+((-\infty, a_\alpha)) + P^+((b_\alpha, +\infty))$. However, $P^+((b_\alpha, +\infty)) = 0$ since the support of P^+ lies at the right-hand side of the core of π. Hence $1 - P^+(\pi_\alpha) = \Pi((-\infty, a_\alpha)) = \alpha$. A similar reasoning holds for P^-, if we notice that $P^-((-\infty, a_\alpha)) = 0$. In fact, we have a more general result:

PROPOSITION 2. *Consider the cumulative distribution function* $F_\lambda = \lambda F^+ + (1 - \lambda)F^-$ *with* $\lambda \in [0, 1]$, *and* P_λ *the associated probability measure. Then* $\forall \lambda \in [0, 1]$, $P_\lambda(\pi_\alpha) = 1 - \alpha$.

Proof. Note that

$$F_\lambda(x) = \begin{cases} \lambda\pi(x) \text{ if } x \leq a_1 \\ \lambda \text{ if } x \in [a_1, b_1] \\ \lambda + (1-\lambda)(1-\pi(x)) \text{ if } x \geq b_1 \end{cases}$$

Now: $P_\lambda(\pi_\alpha) = F_\lambda(b_\alpha) - F_\lambda(a_\alpha) = \lambda + (1-\lambda)(1-\alpha) - \lambda\alpha = 1-\alpha$ □

We also have the following result, laying bare the connection between possibility distributions and the thin clouds of Neumaier,[75] already discussed by Destercke et al.[69]:

PROPOSITION 3. *The set of probability measures for which* $\forall\alpha \in [0,1]$, $P(\pi_\alpha) = 1-\alpha$, *where* π *is the membership function of a fuzzy interval, is* $\mathcal{P}(\pi) \cap \mathcal{P}(1-\pi)$.

Proof. We already know that $\mathcal{P}(\pi) = \{P : \forall\alpha \in [0,1], P(\pi_\alpha) \geq 1-\alpha\}$. Now consider the other inequality $P(\pi_\alpha) \leq 1-\alpha$. Let $\bar{\pi} = 1-\pi$ and note that for continuous membership functions we have that $(\bar{\pi})_\alpha = \overline{\pi_{1-\alpha}}$. Now, $P(\pi_\alpha) \leq 1-\alpha$ is equivalent to $P(\overline{\pi_\alpha}) \geq \alpha$, i.e., $P((\bar{\pi})_{1-\alpha}) \geq \alpha$, or, equivalently, $P((\bar{\pi})_\alpha) \geq 1-\alpha$. So, $\{P : \forall\alpha \in [0,1], P(\pi_\alpha) \leq 1-\alpha\} = \mathcal{P}(1-\pi)$. □

(See Ref. 76 for examples of probability measures whose cumulative distributions lie between F^- and F^+ but are not in the credal set $\mathcal{P}(\pi)$.) Providing a precise description of the content of $\mathcal{P}(\pi)$ is an interesting topic of research.

4.2.5. Possibility Distributions as Likelihood Functions

Another interpretation of numerical possibility distributions is the likelihood function in non-Bayesian statistics (Smets[77], Dubois et al.[78]). In the framework of an estimation problem, the problem is to determine the value of some parameter $\theta \in \Theta$ that characterizes a probability distribution $P(\cdot \mid \theta)$ over U. Suppose that our observations are summarized by the data set $\hat{d}$. The function $P(\hat{d} \mid \theta), \theta \in \Theta$ is not a probability distribution, but a likelihood function $\mathcal{L}(\theta)$: A value a of θ is considered as being all the more plausible as $P(\hat{d} \mid a)$ is higher, and the hypothesis $\theta = a$ will be rejected if $P(\hat{d} \mid a) = 0$ (or is below some relevance threshold). If we extend the likelihood of elementary hypotheses $\lambda(\theta) = cP(\hat{d}|\theta)$ (it is defined up to a positive multiplicative constant c (Ref. 79)), viewed as a representation of uncertainty about θ, to disjunctions of hypotheses, the corresponding set-function Λ should obey the laws of possibility measures[52,80] in the absence of a probabilistic prior, namely, the following properties look reasonable for such a set-function Λ:

- The properties of probability theory enforce $\forall T \subseteq \Theta$, $\Lambda(T) \leq \max_{\theta \in T} \lambda(\theta)$;
- A set-function representing likelihood should be monotonic with respect to inclusion: If $\theta \in T$, $\Lambda(T) \geq \lambda(\theta)$;
- Keeping the same scale as probability functions, we assume $\Lambda(\Theta) = 1$.

Then it is clear that

$$\lambda(\theta) = \frac{P(\hat{d}|\theta)}{\max_{\theta \in \Theta} P(\hat{d}|\theta)},$$

and $\Lambda(T) = \max_{\theta \in T} \lambda(\theta)$, i.e., the extended likelihood function is a possibility measure, and the coefficient c is then fixed. We recover Shafer's proposal of a consonant belief function derived from likelihood information,[36] more recently studied by Aickin.[81] What is interesting to notice is that a conditional probability $P(A \mid B)$ conveys two meanings. It generally represents frequentist information about the frequency of randomly generated objects having property A in class B; conversely, it represents epistemic (nonfrequentist) uncertainty about the class B for an object having property A. It is a bifaced notion with one side that is probabilistic and another side possibilistic. Clearly, acquiring likelihood functions is one way of constructing possibility distributions.

4.3. Possibility Distributions Induced by Human-Originated Estimates

Another source of information for building possibility distributions consists in estimates supplied by human experts on the value of an unknown quantity X of interest, for instance, a failure rate.

4.3.1. Intervals with Confidence Levels

In the most elementary case, such information from a witness or an expert will most naturally take the form of an interval $I = [a, b]$, since we cannot expect precise knowledge generally. A confidence level λ will be attached to this interval, either because the expert expresses some doubts about the estimate, or because the receiver does not fully trust the competence of the expert. This information can be modeled, following Shafer,[36] by a *simple support belief function* with mass $m([a, b]) = \lambda$, while the mass $1 - \lambda$ will be allocated to the widest possible range U for the unknown quantity X, expressing ignorance. Clearly, this procedure yields the hat-shaped possibility distribution π, presented in Equation 1, of the form $\pi(u) = 1$ if $u \in [a, b]$, and $1 - \lambda$ otherwise.

Now the receiver may sometimes find the interval $[a, b]$ too wide to be informative, or, on the contrary, too narrow to be safe enough. It is natural to collect several such human-originated intervals of various sizes and levels of confidence. In contrast with intervals obtained from the imperfect observation of random experiments, intervals coming from one expert will generally be nested, if the latter displays self-consistency. Considering that there is full dependency between these information items (they come from the same person), the collection of nested intervals $I_1 \subseteq \ldots \subseteq I_n$ with confidence levels λ_i can be viewed as a kind of possibilistic knowledge base and correspond to the "double-staircase-shaped" possibility

distribution of Equation 4

$$\pi(u) = \min_{i=1}^{n} \max(I_i(u), 1 - \lambda_i) = \sum_{i:u \in I_i} m(I_i)$$

where $m(I_i) = \lambda_i - \lambda_{i-1}$. Should the pieces of information (I_i, λ_i) come from independent sources, one would be led to replace min by product in this expression (which would be in full agreement with Dempster's rule of combination). However, the intervals would have less chance to be nested.

One may be inspired by the way probability distributions are elicited from experts. In this case, information is requested in the form of quantiles of the distributions, typically, the interval $[a, b]$ is such that $P((-\infty, a]) = 0.05$ and $P([b, +\infty)) = 0.05$. Clearly, the hat-shaped possibility distribution induced by the piece of information $[a, b]$ with confidence 0.1 is a weak form of the information supplied by the two quantiles. This information is sometimes augmented by the 0.5 quantile (the median). In that case, a more faithful representation of this information is in the form of a belief function with disjoint focal sets.

4.3.2. Expert-Originated Statistical Parameters

Another kind of information experts may supply consists of parameters of an otherwise unknown distribution when the unknown quantity is a random variable. In this case, one may use probabilistic inequalities to derive a possibility distribution. For instance, if the expert has a clear idea of the mean $\hat{x}$ of the probability measure P, and of its standard deviation σ, the Chebychev inequality gives us a family of inequalities $P(\overline{A_\lambda}) \leq \min(1, \frac{1}{\lambda^2})$, where $A_\lambda = [\hat{x} - \lambda \cdot \sigma, \hat{x} + \lambda \cdot \sigma]$. This nested family corresponds to the possibility distribution $\pi(\hat{x} - \lambda \cdot \sigma) = \pi(\hat{x} + \lambda \cdot \sigma) = \min(1, \frac{1}{\lambda^2})$.[73] It is consistent with any probability measure with mean $\hat{x}$ and standard deviation σ. The work of Mauris[70] presented above allows to derive a nontrivial possibility distribution from the mere knowledge of the mode of a distribution. Note that the mode corresponds to the idea of most frequently observed values and sounds like a more likely information to be supplied by one expert than for instance the mean value, or even the median. The mode is generally not unique but corresponds to the idea of usual value, whereas the mean value may correspond to seldom observed values, e.g., located between modes. If the information about the mode is supplemented by a safe range for the unknown quantity, the triangular fuzzy number with such mode and support is a faithful representation of this information,[70,73] and it a special case of Gauss inequality,[82] which dates back to 1823; see Baudrit and Dubois[76] for more details on possibility distributions induced by the knowledge of statistical parameters.

4.3.3. From Subjective Probabilities to Subjective Possibilities

One traditional approach to elicitate probability distributions is via fair betting rates. Namely, the subjective probability $P(A)$ of a singular event A, as per an agent,

is viewed as the fair price of a lottery ticket that provides one dollar to this agent if this event occurs. Fairness means that the buyer would accept to sell the lottery ticket at the same price. It is clear that for any k mutually exclusive and exhaustive events $A_1, \ldots, A_k$, we must have that $\sum_{i=1}^{k} P(A_i) = 1$ by fear of losing money otherwise. If there is no reason to consider one event more likely than another then $P(A_i) = 1/k$ for such all events.

The legitimacy of this representation of the epistemic state of an agent has been questioned.[33,36,83] In particular, it can be considered ambiguous. It presupposes a one-to-one function between epistemic states and probability distributions. However, the subjective distribution would be uniform in both cases where the agent is fully ignorant and when he perfectly knows that the stochastic process generating the events is pure randomness. So it is actually a many-to-one mapping, and, given a subjective probability assignment provided by an expert following the betting rate protocol, there is no clue about the precise epistemic state that led to those betting rates.

If we stick to the Bayesian methodology of eliciting fair betting rates from the agent, but we reject the assumption that degrees of beliefs coincide with these betting rates, it follows that the subjective probability distribution supplied by an agent is only a trace of this agent's beliefs. While, in the presence of partial information, beliefs can be more faithfully represented by a set of probabilities, the agent is forced to be additive by the postulates of exchangeable bets. In the transferable belief model,[84] the agent's epistemic state is supposed to be represented by a random epistemic set with mass m, and the subjective probability provided by the Bayesian protocol is called the *pignistic probability*[85] (also known as Shapley value in the game-theoretic literature[86]):

$$pp(u_i) = \sum_{j:u_i \in E_j} \frac{m(E_j)}{|E_j|}. \tag{10}$$

This is an extension of the Laplace principle of insufficient reason, whereby uniform betting rates are assumed inside each focal set. Then, given a subjective probability, the problem consists in reconstructing the underlying belief function.

There are clearly several random sets $\{(E_i, m(E_i)) : i = 1 \ldots n\}$ corresponding to a given pignistic probability. It is in agreement with the minimal specificity principle to consider, by default, the least informative among those. It means adopting a pessimistic view on the agent's knowledge. This is in contrast with the case of statistical probability distributions where the available information consists of observed data. Here, the available information being provided by an agent, it is not assumed that the epistemic state is a unique probability distribution. The most elementary way of comparing belief functions in terms of informativeness consists in comparing contour functions in terms of the specificity ordering of possibility distributions. Dubois et al.[87] proved that the least informative random set with a prescribed pignistic probability $p_i = pp(u_i), i = 1, \ldots, n$ is unique and consonant. It is based on a possibility distribution π^{sub}, previously suggested in Ref. 88 with a

totally different rationale:

$$\pi^{sub}(u_i) = \sum_{j=1}^{n} \min(p_j, p_i). \tag{11}$$

More precisely, let $\mathcal{F}(p)$ be the set of random sets R with pignistic probability p. Let π_R be the possibility distribution induced by R using the one-point coverage Equation (6). Define R_1 to be at least as informative a random set as R_2 whenever $\pi_{R_1} \leq \pi_{R_2}$. Then, the least informative R in $\mathcal{F}(p)$ is precisely the consonant one such that $\pi_R = \pi^{sub}$. Note that, mathematically, Equation 10, when restricted to consonant masses of possibility measures, defines the converse function of Equation 11, i.e., they define a bijection between possibility and probability distributions. Namely, starting from $\pi_1 \geq \ldots \geq \pi_n$ defining the possibility distribution π, computing its associated pignistic probability pp, we have that $\pi^{sub}(u_i) = \sum_{j=1}^{n} \min(pp(u_j), pp(u_i)) = \pi_i$.

By construction, π^{sub} is a subjective possibility distribution. Its merit is that it does not assume human knowledge is precise, like in the subjective probability school. The subjective possibility distribution (11) is less specific than the optimal transformation (7), as expected, i.e., $\pi^{sub} > \pi_p$, generally. The transformation (11) was first proposed in Ref. 88 for objective probability, interpreting the empirical necessity of an event as the sum of excesses of probability of realizations of this event with respect to the probability of the most likely realization of the opposite event.

5. CONCLUSION

One of the most promising seminal off-spring of fuzzy sets introduced in Zadeh's 1965 paper[89] is possibility theory. Possibility theory bridges the gap between artificial intelligence and statistics. The above survey of methods for deriving possibility distributions from data or human knowledge suggests that this framework is one way to go in the problem of membership function assessment. Of course, not all fuzzy sets are possibility distributions, especially those representing utility functions, or those fuzzy sets with a conjunctive interpretation,[2] like a vector of ratings in multifactorial evaluations. However, possibility theory clarifies the role of fuzzy sets in uncertainty management and explains why probability degrees, viewed as frequency or betting rates, can be used to derive membership functions.

References

1. Zadeh LA. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1978; 1: 3–28 (originally, Memo UCB/ERL M77/12, 1977, University of California, Berkeley).
2. Zadeh LA. PRUF - a meaning representation language for natural languages. Int J Man-Mach Stud 1978; 10: 395–460.
3. Bosc P, Pivert O. Fuzzy preference queries to relational databases. London: Imperial College Press; 2012
4. Lodwick W, Kacprzyk J. (Eds.). Fuzzy optimisation: Recent advances and applications. Studies in Fuzziness and Soft Computing, Vol. 254. Berlin: Springer; 2010.

5. Dubois D, Prade H. Fuzzy set and possibility theory-based methods in artificial intelligence. Artif Intell 2003; 148(1–2): 1–9.
6. Dubois D, Kerre E, Mesiar R, Prade H. Fuzzy interval analysis. In: Dubois D, Prade H, editors. Fundamentals of fuzzy sets. The handbooks of fuzzy sets series. Boston, MA: Kluwer; 2000. pp 483–581.
7. Dubois D, Lang J, Prade H, Possibilistic logic. In: Gabbay DM, Hogger CJ, Robinson JA, Nute D, editors. Handbook of logic in artificial intelligence and logic programming, Vol. 3. Oxford, UK: Oxford University Press; 1994. pp 439–513.
8. Dubois D, Prade H. Possibility theory and its applications: where do we stand? In: Kacprzyk J, Pedrycz W, editors. Springer handbook of computational intelligence. Berlin: Springer; 2015. pp 31–60
9. Shackle GLS. The expectational dynamics of the individual. Economica 1943; 10(38): 99–129.
10. Shackle GLS. Expectation in economics. Cambridge, UK: Cambridge University Press; 1949. 2nd ed., 1952.
11. Shackle GLS. Decision, order and time in human affairs (2nd ed.). Cambridge, UK: Cambridge University Press; 1961.
12. Lewis DK. Counterfactuals and comparative possibility. J Philos Logic 1973;2:418–446,. Reprinted in *Ifs*, (W. L. Harper, R. Stalnaker, G. Pearce, eds.), Dordrecht, the Netherlands: D. Reidel; 1981. pp 57–85.
13. Zadeh LA. A note on similarity-based definitions of possibility and probability. Inform Sci 2014; 267: 334–336.
14. Yager RR. An introduction to applications of possibility theory. Human Syst Manag 1983; 3: 246–269.
15. Dubois D, Prade H. Possibility theory. New York: Plenum Press; 1988.
16. Zadeh LA. A theory of approximate reasoning. In: Hayes JE, Mitchie D, Mikulich LI, editors. Machine intelligence, Vol 9. Elsevier: 1979. pp 149–194.
17. Dubois D, Prade H, Smets P. New semantics for quantitative possibility theory. In: Benferhat S, Hunter A, editors. Symbolic and quantitative approaches to reasoning with uncertainty (ECSQARU 2001). LNCS, Vol. 2143. Berlin: Springer-Verlag; 2001. pp 410–421.
18. Dubois D, Prade H. The logical view of conditioning and its application to possibility and evidence theories. Int J Approx Reason 1990; 4(1): 23–46.
19. Dempster AP. Upper and lower probabilities induced by a multivalued mapping. Ann Math Stat 1967; 38: 325–339.
20. Dubois D, Hajek P, Prade H. Knowledge-driven versus data-driven logics. J Logic Lang Inform 2000; 9: 65–89.
21. Dubois D, Prade H. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. Fuzzy Sets Syst 2009; 160(10): 1355–1366.
22. Spohn W. A general, nonprobabilistic theory of inductive reasoning. In: Shachter RD, Levitt TS, Kanal LN, Lemmer JF, editors. Uncertainty in artificial intelligence, Vol 4. Amsterdam: North Holland; 1990. pp 149–158.
23. Grove A. Two modellings for theory change. J Philos Logic 1988; 17: 157–170.
24. Spohn W. Ordinal conditional functions: a dynamic theory of epistemic states. In: Harper WL, Skyrms B, editors. Causation in decision, belief change, and statistics, Vol 2. Norwell, MA: Kluwer; 1988. pp 105–134.
25. Spohn W. The laws of belief: ranking theory and its philosophical applications. Oxford, UK: Oxford University Press; 2012.
26. Benferhat S, Dubois D, Prade H, Williams M-A. A framework for iterated belief revision using possibilistic counterparts to Jeffrey's rule. Fundam Inform 2010; 99(2): 147–168.
27. Gärdenfors P. Knowledge in flux. Cambridge, MA: MIT Press; 1988.
28. Dubois D. Belief structures, possibility theory and decomposable confidence measures on finite sets. Comput Artif Intell (Bratislava) 1986; 5(5): 403–416.
29. Shafer G. Belief functions and possibility measures. In: Bezdek JC, editor. Analysis of fuzzy information, Vol I: Mathematics and logic. Boca Raton, FL: CRC Press; 1987. pp 51-84.

30. Dubois D, Prade H. Fuzzy sets and systems: Theory and applications. New York: Academic Press; 1980.
31. Dubois D, Prade H. When upper probabilities are possibility measures. Fuzzy Sets Syst 1992; 49: 65–74.
32. De Cooman D, Aeyels D. Supremum-preserving upper probabilities. Inform Sci 1999; 118: 173–212.
33. Walley P. Statistical reasoning with imprecise probabilitiess. London: Chapman and Hall; 1991.
34. Giles R. Foundations for a theory of possibility. In: Gupta M, Sanchez E, editors. Fuzzy information and decision processes. Amsterdam: North-Holland; 1982. pp 183–196.
35. Walley P, De Cooman G. A behavioural model for linguistic uncertainty. Inform Sci 1999; 134: 1–37.
36. Shafer G. A Mathematical theory of evidence. Princeton, NJ: Princeton University Press; 1976.
37. Dubois D, Prade H. On several representations of an uncertain body of evidence. In: Gupta M, Sanchez E, editors. Fuzzy information and decision processes. Amsterdam: North-Holland; 1982. pp 167–181.
38. Dubois D, Prade H. Epistemic entrenchment and possibilistic logic. Artif Intell 1991; 50: 223–239.
39. Dubois D, Fargier H, Prade H. Ordinal and probabilistic representations of acceptance. J. Artif Intell Res 2004; 22: 23–56.
40. Dubois D, Fargier H. Prade H. Possibility theory in constraint satisfaction problems: handling priority, preference and uncertainty. Appl Intell 1996; 6: 287–309.
41. Schockaert S, Prade H. Solving conflicts in information merging by a flexible interpretation of atomic propositions. Artif Intell 2011; 175(11): 1815—1855.
42. Dupin de Saint Cyr F, Lang J, Schiex Th. Penalty logic and its link with Dempster–Shafer theory. In: Lopez de Mantaras R, Poole D, editors. Proc Annual Conf on Uncertainty in Artificial Intelligence (UAI'94), Seattle, July 29–31, 1994. San Francisco, CA: Morgan Kaufmann; 1994. pp 204–211.
43. Mamdani EH. Advances in the linguistic synthesis of fuzzy controllers. Int J Man-Mach Stud 1976; 8: 669–678.
44. Benferhat S, Dubois D, Prade H. Practical handling of exception-tainted rules and independence information in possibilistic logic. Appl Intell 1998; 9(2): 101–127.
45. Pearl J. System Z: a natural ordering of defaults with tractable applications to nonmonotonic reasoning. In: Parikh R, editor. Proc 3rd Conf on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove, CA. San Francisco, CA: Morgan Kaufmann; March 4–7, 1990. pp 121–135.
46. Goldszmidt M, Pearl J. Qualitative probability for default reasoning, belief revision and causal modeling. Artif Intell 1996; 84: 52–112.
47. Sudkamp T. Similarity and the measurement of possibility. Actes Rencontres Francophones sur la Logique Floue et ses Applications. Toulouse, France: Cepadues Editions; 2002. pp 13–26.
48. Ruspini EH. On the semantics of fuzzy logic. Int J Approx Reason 1991; 5(1): 45–88.
49. Türksen IB, Bilgic T. Measurement of membership functions: theoretical and empirical work. In: Dubois D, Prade H, editors. Fundamentals of fuzzy sets. The Handbooks of Fuzzy Sets. Norwell, MA: Kluwer; 2000. pp 195–230.
50. Marchant T. The measurement of membership by comparisons. Fuzzy Sets Syst 2004; 148: 157–177.
51. Marchant T. The measurement of membership by subjective ratio estimation. Fuzzy Sets Syst 2004; 148: 179–199.
52. Dubois D. Possibility theory and statistical reasoning. Comput Stat Data Anal 2006; 51: 47–69.
53. Mauris G. Possibility distributions: a unified representation of usual direct-probability-based parameter estimation methods. Int J Approx Reason 2001; 52(9): 1232–1242.

54. Mauris G. A review of relationships between possibility and probability representations of uncertainty in measurement. IEEE Trans Instrum Meas 2013; 62(3): 622–632.
55. Couso I, Dubois D. Statistical reasoning with set-valued information: Ontic vs. epistemic views. Int J Approx Reason 2014;55(7):1502–1518. Special issue on "Harnessing the information contained in low-quality data sources."
56. Ferson S, Ginzburg L, Kreinovich V, Longpré L, Aviles M. Exact bounds on finite populations of interval data. Reliable Comput 2005;11(3):207–233
57. Joslyn C. Measurement of possibilistic histograms from interval data. Int J Gen Syst 1997; 26: 9–33.
58. Couso I, Dubois D, Sanchez L. Random sets and random fuzzy sets as ill-perceived random variables. Springer Briefs in Computational Intelligence. Berlin: Springer; 2014.
59. Ferson S, Kreinovich V, Ginzburg L, Myers DS, Sentz K. Constructing probability boxes and Dempster–Shafer structures. Technical Report SAND2002-4015. Albuquerque, NM: Sandia National Laboratories; 2003.
60. Kampé de Fériet J. Interpretation of membership functions of fuzzy sets in terms of plausibility and belief. In: Gupta M, Sanchez E, editors. Fuzzy information and decision processes. Amsterdam: North-Holland; 1982. pp 93–98.
61. Goodman IR. Fuzzy sets as equivalence classes of random sets. In Yager R, editor. Fuzzy sets and possibility theory: Recent developments. Oxford, UK: Pergamon Press; 1981. pp 327–342.
62. Dubois D, Prade H. Consonant approximations of belief functions. Int J Approx Reason 1990; 4: 419–449.
63. Dubois D, Prade H. Fuzzy sets and statistical data, Eur J Oper Res 1986;25:345–356.
64. Delgado M, Moral S, On the concept of possibility-probability consistency. Fuzzy Sets Syst 1987;21:311–318.
65. Dubois D, Huellermeier E. Comparing probability measures using possibility theory: a notion of relative peakedness. Int J Approx Reason 2007; 45: 364–385.
66. de Campos L, Huete J, Moral S. Probability intervals: a tool for uncertain reasoning. Int J Uncertain Fuzziness Knowl-Based Syst 1994; 2: 167–196.
67. de Campos L, Huete J. Measurement of possibility distributions. Int J Gen Syst 2001; 30(3): 309–346.
68. Masson M, Denoeux T. Inferring a possibility distribution from empirical data. Fuzzy sets Syst 2006; 157: 319–340.
69. Destercke S, Dubois D, Chojnacki E. Unifying practical uncertainty representations, part II: clouds. Int J Approx Reason 2008; 49: 664–677.
70. Mauris G. Inferring a possibility distribution from very few measurements. In: Dubois D, et al., editors. Soft methods for handling variability and imprecision (SMPS 2008). Advances in Soft Computing, Vol 48. Berlin: Springer; 2008. pp 92–99.
71. Machol RE, Rosenblatt J. Confidence intervals based on a single observation. Proc IEEE 1966; 54: 1087–1088.
72. Dubois D, Prade H, Sandri SA. On possibilityprobability transformations. In: Lowen R, Lowen M, editors. Fuzzy Logic: State of the Art, Kluwer Academic Publ.; 1993, pp 103–112.
73. Dubois D, Foulloy L, Mauris G, Prade H. Possibility/probability transformations, triangular fuzzy sets, and probabilistic inequalities. Reliable Comput 2004; 10: 273–297.
74. Couso I, Montes S, Gil P. The necessity of the strong α-cuts of a fuzzy set. Int J Uncertain Fuzziness Knowl-Based Syst 2001; 9(2): 249–262.
75. Neumaier A. Clouds, fuzzy sets and probability intervals. Reliable Comput 2004; 10: 249–272.
76. Baudrit C and Dubois D. Practical representations of incomplete probabilistic knowledge. Comput Stat Data Anal 2006; 51: 86–108.
77. Smets P. Possibilistic inference from statistical data. In: Ballester A, editor. Proc Second World Conference on Mathematics at the Service of Man, Las Palmas, Spain;1982. pp 611–613.

78. Dubois D, Moral S, Prade H. A semantics for possibility theory based on likelihoods. J Math Anal Appl 1997; 205: 359–380.
79. Edwards WF. Likelihood. Cambridge, UK: Cambridge University Press; 1972.
80. Coletti G, Scozzafava R. Coherent conditional probability as a measure of uncertainty of the relevant conditioning events. In: Proc 7th Eur Conf on Symbolic and Quantitative Approaches to Reasoning under Uncertainty, Aalborg, Denmark, July 2–5, 2003 (ECSQARU03). LNAI 2711. Berlin: Springer-Verlag; 2003. pp 407–418.
81. Aickin M. Connecting Dempster–Shafer belief functions with likelihood-based inference. Synthese 2000l; 123(3): 347–364.
82. Upton G, Cook I. Gauss inequality. In: A dictionary of statistics. Oxford, UK: Oxford University Press; 2008.
83. Dubois D, Prade H. Formal representations of uncertainty. In: Decision-making Process-Concepts and Methods. In: Bouyssou D, Dubois D, Pirlot M, Prade H, editors. London: ISTE; 2009. Vol 3, pp 85–156.
84. Smets P, Kennes R. The transferable belief model. Artif Intell 1994; 66: 191–234.
85. Smets P. Constructing the pignistic probability function in a context of uncertainty, In Henrion M, et al., editors. Uncertainty in artificial intelligence, Vol 5, Amsterdam: North-Holland; 1990. pp 29–39.
86. Shapley S. A value for *n*-person games. In Kuhn HW, Tucker AW, editors. Contributions to the theory of games, Vol II. Princeton, NJ: Princeton University Press; 1953. pp 307–317.
87. Dubois D, Prade H, Smets P. A definition of subjective possibility. Int J Approx Reason 2008; 48: 352–364.
88. Dubois D, Prade H. Unfair coins and necessity measures: towards a possibilistic interpretation of histograms. Fuzzy Sets Syst 1983; 10: 15–20.
89. Zadeh LA. Fuzzy sets. Inform Control 1965; 8: 338–353.