A Survey on Quantitative Possibility Theory in Artificial Intelligence: A Convenient Epistemic Uncertainty Model

Sébastien Destercke[1] , **Didier Dubois**[2] , **Henri Prade**[2]
[1]Heudiasyc, Univ. of Technology of Compiègne, France
[2]IRIT, Univ. of Toulouse, France
sebastien.destercke@cnrs.fr, {didier.dubois, henri.prade}@irit.fr

Abstract

Quantitative (or numerical) possibility theory offers a simple but yet very expressive setting for handling epistemic uncertainty and in particular imprecise probabilities. The paper surveys the basic ideas and notions underlying numerical possibility theory, its relation to other uncertainty settings and its use in AI-related issues. Numerical possibility theory looks of interest for coping with imperfect statistical information, especially non-Bayesian statistics relying on likelihood functions and confidence intervals. Quantitative possibility theory can be used in inference, machine learning, tracking and information fusion, and finally preference modeling.

1 Introduction

Modeling incomplete information and partial ignorance has been an important issue in AI for about thirty-five years. This covers a large territory that ranges from reasoning under incomplete information and nonmonotonic reasoning to the handling of epistemic uncertainty [Dubois *et al.*, 1996]. The accurate representation of epistemic uncertainty in case of partial ignorance requires two numbers for quantifying belief and plausibility, as in possibility theory [Dubois and Prade, 1988], evidence theory [Shafer, 1976], or upper and lower probabilities [Walley, 1991].

Unlike the last two frameworks, possibility theory may be qualitative or quantitative [Dubois and Prade, 1998] depending if we use an ordinal setting for ranking propositions in terms of their certainty of being true, or a quantitative scale. Mathematically speaking, quantitative possibility theory can be viewed as a special case of evidence theory, or of upper and lower probabilities. It was recognized early on as the simplest framework for effectively representing imprecise probabilities [Walley, 1996]. Qualitative possibility theory and its close relationship to classical logic has been recently surveyed in [Dubois and Prade, 2025]. This article aims to supplement the previous article with a study on the quantitative aspects of possibility theory and their relationship to probability theory. This paper revisits and updates previous surveys such as [Dubois, 2006].

The paper is organized as follows. Section 2 reviews the basics of quantitative possibility theory by presenting its set functions, distinguishing it from qualitative possibility theory. We recall two types of conditioning that account for the ideas of a shift in reference class and a revision of beliefs, respectively. Then we survey possibilistic Bayesian networks and the underlying notion of independence. Section 3 is devoted to the relations with other uncertainty settings; we first interpret possibility distributions as likelihood functions in the absence of prior information; next, the relationship between possibility theory and evidence theory, as well as with imprecise probabilities and the representation of probability families using possibility distributions, are examined; then we present two transformations between probability and possibility serving different purposes; the section ends with so-called p-boxes and clouds, where the use of pairs of possibility distributions provides a more accurate representation of probability families. Finally, we recall the view of possibility as infinitesimal probabilities. In each case the expressive power of possibility theory is emphasized. Section 4 surveys different uses of possibility theory in uncertainty propagation, statistical inference, machine learning, tracking and information fusion, and finally preference modeling.

2 Basics of Quantitative Possibility Theory

Possibility theory is a relatively recent uncertainty theory since the term "theory of possibility" was coined just under 50 years ago by L. A. Zadeh [1978], when he defined a possibility measure from a fuzzy set membership function for evaluating queries relative to natural language statements such as "what is the possibility that John is more than 25 given that John is young?" (where "young" is represented by a fuzzy set of ages). However an epistemic notion of possibility already appeared in the calculus of degrees of "potential surprise" proposed by the economist G. L. S. Shackle [1949] (potential surprise being evaluated as a degree of impossibility). The philosopher Isaac Levi was the first to relate Shackle's ideas to probability theory [Levi, 1967]. First, we review the concept of possibility distribution and the associated set functions in possibility theory, before introducing definitions of conditioning for quantitative possibility theory. ‘

2.1 Possibility Distribution and Set Functions

Possibility Distribution The basic building block of possibility theory is the notion of possibility distribution, supposed to represent the available information about the value of a single-valued quantity of interest. A possibility distribution is a mapping π_x from a referential U to a scale $\mathcal{S}$ and $\pi_x(u) \in \mathcal{S}$ represents the degree of possibility that u is the actual value of the quantity x. The scale $\mathcal{S}$ is usually assumed to be bounded and linearly ordered. In quantitative possibility theory, $\mathcal{S}$ is just the unit interval $[0, 1]$. $\pi_x(u) = 0$ means that $u \in U$ is a value fully impossible for x, while $\pi_x(u) = 1$ means that u is totally possible ; note that distinct values may have a degree of possibility equal to 1 (a real interval is a special case of a possibility distribution on $\mathbb{R}$).

The normalization condition $\exists u \in U, \pi_x(u) = 1$ ensures that at least one u is fully possible for x; it expresses the consistency of the epistemic state described by π_x. The state of vacuous information about x is represented by the distribution $\pi_x = \pi^\top$ such that $\forall u \in U, \pi^\top(u) = 1$, i.e., any value is totally possible for x. In contrast, complete information $x = u_0$ is represented by $\pi_x(u_0) = 1$ and $\forall u \neq u_0, \pi_x(u) = 0$.

Specificity A possibility distribution π is said to be at least as specific as another π' if and only if $\forall u \in U, \pi(u) \leq \pi'(u)$ [Yager, 1983]. A non-specificity index having entropy-like properties has been proposed in [Higashi and Klir, 1982], from which informational distances can be defined; see [Ramer, 1990] for properties and characterizations. Possibility theory is governed by the *principle of minimal specificity* that is a principle of least commitment that assigns the highest degree of possibility to each u compatible with the available pieces of information.

Possibility and Necessity Two dual set functions, a possibility measure Π and a necessity measure N such that $N(A) = 1 - \Pi(A^c)$ for any event $A \subseteq U$ and its complement $A^c = U \setminus A$, are induced by a possibility distribution π: $\Pi(A) = \sup_{u \in A} \pi_x(u); N(A) = 1 - \Pi(A^c) = \inf_{u \notin A} 1 - \pi_x(u)$ Due to duality, $N(A)$ is all the higher ($x \in A$ is all the more certain) as the opposite event is less possible. $\Pi(A)$ (resp. $N(A)$) evaluates to what extent event $x \in A$ is consistent with π_x (resp. A can be deduced from π_x). By convention $\Pi(\emptyset) = 0$, thus $N(U) = 1$, and by normalization, $\Pi(U) = 1$, thus $N(\emptyset) = 0$.

Possibility measures are characterized by the "maxitivity" property $\forall A, B, \Pi(A \cup B) = \max(\Pi(A), \Pi(B))$, and necessity measures are "minitive":$\forall A, B, N(A \cap B) = \min(N(A), N(B))$. Due to the normalization of π, $\min(N(A), N(A^c)) = 0$ and $\max(\Pi(A), \Pi(A^c)) = 1$, or equivalently $\Pi(A) = 1$ whenever $N(A) > 0$, namely something somewhat certain should be first fully possible. While a probability P is self-dual, and such that $P(A^c) = 0 \Rightarrow P(A) = 1$, possibility and necessity are such that $N(A^c) = 0 \not\Rightarrow N(A) = 1$ (but $\Pi(A^c) = 0 \Rightarrow \Pi(A) = 1$)

A possibility distribution π_x can represent a finite family of nested confidence subsets $\{A_1, A_2, \cdots, A_m\}$ of U where $A_i \subset A_{i+1}$, $i = 1, \cdots, m-1$. Each subset A_i is attached a positive confidence level λ_i. The links between the confidence levels λ_i's and the degrees of possibility are defined by postulating λ_i is the degree of necessity (i.e., certainty) $N(A_i)$ of A_i. It entails that $\lambda_1 \leq \cdots \leq \lambda_m$ due to the monotonicity of the necessity function N. The possibility distribution equivalent to the weighted family $\{(A_1, \lambda_1), (A_2, \lambda_2), \cdots, (A_m, \lambda_m)\}$ is defined as the least informative / specific possibility distribution π that obeys the constraints $\lambda_i \leq N(A_i), i = 1, ..., m$.

Combination, Projection A joint possibility distribution $\pi_{x_1,\cdots,x_n}$ is a fuzzy restriction on the simultaneous values of several variables $x_1, \cdots, x_n$ on a Cartesian product $U_1 \times \cdots \times U_n$. The projection on U_i is nothing but the possibility of the event $x_i = u_i$ the other variables being unrestricted: $\pi_{x_i}(u_i) = \sup_{u_j, j \neq i} \pi_{x_1,\cdots,x_n}(u_1, \cdots, u_n)$. Generally, $\pi_{x_1,\cdots,x_n} \leq \min(\pi_{x_1}, \cdots, \pi_{x_n})$. When we consider two variables x and y with possibility distributions π_x and π_y, then by virtue of the principle of minimal specificity, the joint distribution is defined as $\pi_{x,y} = \min(\pi_x, \pi_y)$. This mechanism of combination/projection is the basis of Zadeh's theory of approximate reasoning [1979] and his extension principle.

Guaranteed Possibility Apart from Π and N, two other quantities can be associated to a possibility distribution π_x (see e.g., [Benferhat *et al.*, 2002]): the guaranteed possibility $\Delta(A) = \inf_{u \in A} \pi_x(u)$ and the dual set function $\nabla(A) = 1 - \Delta(A^c) = \sup_{u \notin A} 1 - \pi_x(u)$. There is no link between $\Delta(A)$ and $N(A)$. $\Delta(A)$ estimates to what extent *all* the values in A are actually possible for x; clearly $\Delta(A) \leq \Pi(A)$. $\nabla(A) \geq N(A)$ estimates a potential certainty (a value u in A^c with a low degree of possibility is only a necessary condition for having $x \in A$ somewhat certain). Δ and ∇ are decreasing set functions and obey a maximal specificity principle: indeed a set of constraints $\Delta(A_j) \geq \rho_j, j = 1, \cdots, k$ induces a lower bound on $\pi_x(u) \geq \max_{j=1,k} \delta_{(A_j,\rho_j)}$ where $\delta_{(A_j,\rho_j)}(u) = \rho_j$ if $u \in A_j$ and $\delta_{(A_j,\rho_j)}(u) = 0$ otherwise.

Quantitative vs. Qualitative Possibility As can be seen the four set functions of possibility theory do not require a numerical scale to be defined since sup and inf make sense as well on a finite ordered scale (and $1 - \cdot$ is then just an order-reversing map of the scale). Thus possibility theory is a two-faced setting. Due to the qualitative nature of the operations defining its set functions, possibility theory fits very well with logic, see the survey [Dubois and Prade, 2025].

However using a numerical scale facilitates its comparison with probability theory, and lays bare links between them. It opens the way to other types of applications surveyed in this paper. When it comes to defining conditioning the options diverge between the qualitative and the quantitative setting as we are going to see.

2.2 Conditioning

The first notion of conditional possibility measure was qualitative and obeyed the equation [Hisdal, 1978]:

$\forall B, B \cap A \neq \emptyset, \Pi(A \cap B) = \min(\Pi(B \mid A), \Pi(A))$.

In order to avoid discontinuity problems in the numerical case, min is replaced by the product, which comes close to probability. Then $\Pi(B \mid A) = \frac{\Pi(A \cap B)}{\Pi(A)}$ if $\Pi(A) \neq 0$ and $N(B \mid A) = 1 - \Pi(B^c \mid A)$ by duality. When conditioning is defined in the numerical setting via the product rule, then the product-based symmetric independence notion $\Pi(A \cap B) = \Pi(A) \cdot \Pi(B)$ is trivially equivalent to $\Pi(B \mid A) = \Pi(B)$ for positive values and is close to probabilistic independence.

Possibilistic Bayes theorem looks like the probabilistic one: $\Pi(B \mid A) \cdot \Pi(A) = \Pi(A \mid B) \cdot \Pi(B)$, but departs from it due to maxitivity of possibility measures. Indeed, consider the problem of testing hypothesis H against its complement, upon observing evidence E; applying the formula of total possibility $\Pi(E) = \max(\Pi(E \mid H) \cdot \Pi(H), \Pi(E \mid H^c) \cdot \Pi(H^c))$, we get $\Pi(H \mid E) = \min(1, \frac{\Pi(E \mid H)\cdot\Pi(H)}{\Pi(E \mid H^c)\cdot\Pi(H^c)})$. In case of lack of prior information, the uniform possibility $\Pi(H) = \Pi(H^c) = 1$ is used, which leads to $\Pi(H \mid E) = \min(1, \frac{\Pi(E \mid H)}{\Pi(E \mid H^c)})$ and $\Pi(H^c \mid E)$ computed likewise. It comes down to comparing $\Pi(E \mid H)$ and $\Pi(E \mid H^c)$, which is similar to likelihood ratio tests (the likelihood function is renormalized via a proportional rescaling [Edwards, 1972]). This approach has been successfully used in practice [Lapointe and Bobée, 2000]. Besides, possibilistic counterparts to Jeffrey's rule of probabilistic conditioning for revision under uncertain inputs exist [Benferhat *et al.*, 2010].

There is another type of possibilistic conditioning that echoes the following writing of conditional probability: $P(B \mid A) = \frac{P(B\cap A)}{P(A\cap B)+1-P(A^c\cup B)}$; it is defined by $\Pi(B \mid_f A) = \frac{\Pi(B\cap A)}{\Pi(A\cap B)+1-\Pi(A^c\cup B)} = \frac{\Pi(B\cap A)}{\Pi(A\cap B)+N(A\cap B^c)}$ and by $N(B \mid_f A) = 1 - \Pi(B \mid_f A) = \frac{N(A\cap B)}{N(A\cap B)+1-\Pi(A\cap B^c)}$. It can be shown that $\Pi(B \mid_f A)$ (resp. $N(B \mid_f A)$) is a possibility (resp. necessity) measure: indeed the set function is maxitive (resp. minitive) [Dubois and Prade, 1997]. This conditioning makes sense when possibility measures are viewed as upper bounds of ill-known probabilities; see subsection 3.3.

2.3 Possibilistic Bayesian Nets

Similarly to joint probability distributions, a joint possibility distribution associated with ordered variables $X_1, \ldots, X_n$ can be decomposed in terms of conditional possibility distributions using the chain rule:
$\pi(X_1,..., X_n) = \pi(X_n \mid X_1,..., X_{n-1}) \cdot \ldots \cdot \pi(X_2 \mid X_1) \cdot \pi(X_1)$
based on product (in the qualitative setting the product is replaced by $\min$). See [Fonck, 1997] for a thorough study on conditional independence between sets of variables and the graphoid structure underlying possibilistic independence. A comparison of the inference mechanisms in the probabilistic and possibilistic frameworks can be found in [Ayachi *et al.*, 2014] inside a generic setting for compilation-based inference. The handling of uncertain inputs in possibilistic networks is addressed in [Benferhat and Tabia, 2014]. The product-based merging of possibilistic networks with the same DAG structure can be performed in polynomial time; solutions to merge possibilistic networks with different structures have been also proposed [Benferhat and Titouna, 2009]. Links between product-based possibilistic Bayesian nets and penalty logic and a product-based quantitative possibilistic logic have been established [Khellaf-Haned and Benferhat, 2014]. Such translations are interesting for taking advantage of each format at the inferential level. See also [Boutouhami *et al.*, 2017] for the handling of uncertainty in ontologies based on product-based possibility theory.

Belief revision in possibilistic graphs can be performed based on new observations; the detection of causality is based on so-called interventions, using an adaptation of Pearl's 'do' operator to the possibilistic setting [Benferhat and Tabia, 2010; Benferhat and Smaoui, 2011].

The practical implementation of possibilistic networks requires the specification of conditional possibility tables. To simplify this task, possibilistic counterparts of the noisy probabilistic connectives have been developed. An illustration of the interest for modeling epistemic uncertainty on a human geography modeling problem can be found in [Dubois *et al.*, 2017]. Lastly let us mention the extension of possibilistic networks to interval-valued possibility for handling multiple expert knowledge by compactly encoding families of standard joint possibility distributions [Benferhat *et al.*, 2014].

3 Relation to Other Uncertainty Settings

Quantitative possibility theory, as a simple but expressive representation framework, has multiple remarkable connections with other uncertainty settings, which we review now.

3.1 Likelihood Function as Possibility Distribution

One understanding of a numerical possibility distribution [Dubois and Prade, 2016] is a likelihood function in non-Bayesian statistics, as first suggested in [Shafer, 1976], who calls it "the contour function of a consonant belief function"; see also [Smets, 1982]. A likelihood function $\lambda(\theta) = P(d|\theta)$, is a representation of uncertainty about θ, where $P(\cdot|\theta)$ is a statistical model with parameter θ, and d is an observed dataset. Likelihood functions are defined up to a multiplicative constant, and behave like *relative* possibility distributions. The likelihood interpretation of quantitative possibility theory takes advantage of the probabilistic inequalities $\inf\{P(A \mid b), b \in B\} \leq P(A \mid B) \leq \sup\{P(A \mid b), b \in B\}$ [Dubois *et al.*, 1997] (i.e., a bracketing by set functions Δ and Π). [Denoeux, 2014] adopts the same approach in order to handle low-quality data.

3.2 Possibility Theory and Evidence Theory

When the focal elements of a mass function in evidence theory are nested, the belief function and the plausibility function are said to be consonant and reduce to a necessity and a possibility measure respectively [Shafer, 1987]. The contour function (plausibility of singletons) is then the underlying possibility distribution. Moreover the commonality function reduces to a guaranteed possibility function. Viewing mass functions as random sets, an inclusion relation can be defined between them. It reduces, in case of consonance, to fuzzy set inclusion that underlies specificity [Dubois and Prade, 1986].

Conditioning a quantitative possibility measure as in section 2.2 corresponds exactly to Dempster rule of conditioning in evidence theory. However the Dempster rule of combination does not preserve consonance. This combination applied to two consonant belief functions yields a belief function whose contour function is the product-based fusion of the two corresponding possibility distributions. See [Dubois *et al.*, 2016] for a comparative study of the fusion modes in the different uncertainty settings.

An important class of belief functions is that of separable functions, obtained by applying Dempster combination to non-dogmatic possibility measures (known as simple support functions) that have only two focal sets (including the reference set). A separable function is characterized by a set of weights, known as diffidence weights, that characterize the simple support functions from which it is constructed. A measure of necessity is an example of a separable function. This representation provides an informational order among separable belief functions which, when they are consonant, differs from the possibilistic specificity order [Denoeux, 2008]. In other words, the approach via diffidence weights leads to a branch of the possibility theory where the order of specificity no longer makes sense. It is replaced by an order in terms of possibility ratios, in accordance with likelihood functions [Dubois *et al.*, 2020].

3.3 Possibility as Upper Probability

Zadeh [1978]'s consistency principle claims that what is probable must be possible, which states that the possibility $\Pi(A)$ of an event should be greater than (or equal to) its probability $P(A)$: namely $\forall A, P(A) \leq \Pi(A)$.

In quantitative possibility theory, any interval bounded by dual necessity / possibility measures $[N, \Pi]$ can be interpreted as upper and lower probabilities induced from a specific convex set of probability measures. Let π be a possibility distribution inducing a pair of functions (N, Π). We define the probability family $\mathcal{P}(\pi) = \{P, \forall A \text{ measurable}, P(A) \leq \Pi(A)\}$. In this case, we have $\sup_{P \in \mathcal{P}(\pi)} P(A) = \Pi(A)$ and $\inf_{P \in \mathcal{P}(\pi)} P(A) = N(A)$ [Dubois and Prade, 1992].

The pairs (interval A_i, necessity weight λ_i) - mentioned in Section 2.1, also modeled by simple support functions - can be interpreted as stating that the probability $P(A_i)$ is at least equal to λ_i where A_i is a measurable set (e.g., a confidence interval containing the value of interest). These intervals can be obtained for instance in terms of inter-fractiles of a probability distribution. The corresponding probability family is:

$$\mathcal{P} = \{P, \forall A_i, \lambda_i \leq P(A_i)\}. \tag{1}$$

If the sets A_i are nested ($A_1 \subset A_2 \subset \cdots \subset A_n$), as can be expected for a family of confidence intervals, then P generates upper and lower probabilities of events that coincide with the possibility and necessity measures induced by the least specific possibility distribution that obeys the constraints $\lambda_i \leq N(A_i)$ [Dubois and Prade, 1992] (see the subpart "Possibility and Necessity" in Section 2.1). Note that such a property is characteristic of probability sets induced by possibility distributions: any probability set that can be expressed through Eq. (1) is induced by a possibility distribution, and any (normalised) possibility distribution induces a probability set described as in Eq. (1).

It should be noted that those families are special, as for instance they always contain the probability $p(u) = 1$ for any element $u \in U$ with $\pi_x(u) = 1$. Also, since $\Pi(A) < 1 \implies N(A) = 0$, the interval $[N, \Pi]$ is either of the form $[0, \alpha]$ or $[\beta, 1]$ and always contains a trivial bound. This means that possibility theory cannot express precise probabilities as a special case, nor can precise probabilities express possibilities. As we shall see in Section 4, possibility theory can still be applied to a lot of settings with success, despite this limited expressiveness. In fact, one could argue that this is precisely this simplicity that makes possibility theory an attractive alternative to probability theory in case of incomplete or imprecise information, as it often allows one to derive computationally efficient solutions.

Viewing possibility degrees as upper bounds of probabilities leads to the justification of the $|_f$- conditionalization of possibility measures (see Section 2.2). This view of conditioning can be called 'Bayesian possibilistic conditioning' [Dubois and Prade, 1997; Walley, 1996] in accordance with imprecise probabilities since $\Pi(B|_f A) = \sup\{P(B|A) : P(A) > 0, P \leq \Pi\}$. Such Bayesian possibilistic conditioning contrasts with product-based conditioning which always supplies more specific results than the above.

3.4 Transformations Probability-Possibility

There are two main probability-possibility transformations that serve different purposes and obey different principles [Dubois *et al.*, 1993].

The first one goes from probability to possibility, and is based on the fact that a possibility measure encodes a family of probability measures, see Section 3.3. The goal is to transform a frequentist probability distribution into a maximally specific possibility distribution whose associated set of probability measures contains the former. Since a frequentist probability distribution contains all the information gathered from the observation of a random phenomenon, maximizing specificity intends to preserve the original information as much as possible [Dubois *et al.*, 2004].

The possibility distribution π obtained from a probability distribution p should thus satisfy
- $P(A) \leq \Pi(A), \forall A \subseteq U$
- p and π are order-equivalent:$\forall u, p(u) > p(u') \Leftrightarrow \pi(u) > \pi(u')$
- π is maximally specific (any other solution π' is s.t. $\pi \leq \pi'$)

In the finite case $U = \{u_1, u_2, \cdots, u_n\}$, given a probability distribution p with probability values $p_1 \geq p_2 \geq \cdots \geq p_n$ ($p_i = p(u_i)$) the solution $\pi = (\pi_1, \pi_2, \cdots, \pi_n)$ with $\pi_i = \pi(u_i)$ is unique and given by $\pi_1 = 1$, and for $i > 1$:

$$\pi_i = \left| \begin{array}{l} \Sigma_{j=i,n} p_j \text{ if } p_{i-1} > p_i \\ \quad \pi_{i-1} \text{ otherwise} \end{array} \right.$$

In the continuous real case (for unimodal probability densities), the obtained possibility distribution encodes a nested family of tightest dispersion intervals, with probabilistic guarantees, around the mode of the statistical distribution considered, which is much more expressive than a simple variance. See [Dubois *et al.*, 2004; Mauris, 2011] for more references.

There is a close connection between the ordering on probability distributions in terms of relative "peakedness" and the standard specificity ordering on possibility distributions that can be derived by means of the above probability–possibility transformation [Dubois and Hüller-meier, 2007].

The second transformation [Dubois *et al.*, 1993] goes from possibility to probability and obeys a least commitment principle, i.e., the third requirement of the first transformation is replaced by
- p contains as much uncertainty as possible.
while the two first requirements are kept.

In the finite case, it is defined as $p_i = \Sigma_{j=i}^{n} \frac{1}{j}(\pi_j - \pi_{j+1})$. where $\pi_1 \geq \cdots \geq \pi_n \geq \pi_{n+1} = 0$. This probability is the gravity center of the set $\mathcal{P} = \{P \mid \forall A, P(A) \leq \Pi(A)\}$ of probability distributions dominated by Π. It is also known as a special case of Smets' *pignistic* transform, which turns a mass function from evidence theory into a probability distribution for decision purposes [Dubois *et al.*, 2004].

The converse mapping turns a probability distribution into a possibility distribution, given by:

$\pi_i = \Sigma_{j=1}^n \min(p_j, p_i), i = 1, \ldots, n$

It preserves the shape of the distributions ($\pi_i > \pi_{i+1} \Leftrightarrow p_i > p_{i+1}$), but the obtained possibility distribution is less specific than the one obtained with the first transformation above. Interestingly, it is the least informative belief function having a given pignistic probability: it is is unique and consonant [Dubois *et al.*, 2008]. This transform thus fits with a subjective view of probability.

3.5 P-boxes and Clouds

A possibility distribution is a very compact representation of (some types of) imprecise probability. However, as already said probability intervals induced by such a possibility distribution are of the form $[\beta, 1]$ or $[0, \alpha]$, which points out a lack of expressiveness. One way to improve the latter while staying compact is to use two possibility distributions (π^+, π^-) instead of one and consider the convex set of probabilities $\mathcal{P}(\pi^+) \cap \mathcal{P}(\pi^-)$. The probability of an event A lies inside the interval $[\max(N^+(A), N^-(A)), \min(\Pi^+(A), \Pi^-(A))]$ (when not empty). There are several examples of such representations (see [Destercke *et al.*, 2008] for details):

- *P-boxes* are ordered pairs $(F^+ \geq F^-)$ of cumulative probability distributions that bracket an ill-known one. Note that cumulative probability distributions can be viewed as possibility distributions induced by nested intervals of the form $(a, +\infty)$. P-boxes are special cases of belief functions.

- *Clouds*, proposed by Neumaier, are defined by two distributions π, δ where $\delta \leq \pi$, where $\pi^+ = \pi$ and $\pi^- = 1 - \delta$ are possibility distributions. The probability bounds obtained correspond to belief and plausibility measures, provided that π and δ are comonotonic. P-boxes are special cases of comonotonic clouds.

3.6 Possibility as Infinitesimal Probability

Ranking functions [Spohn, 2012] introduce an implausibility ranking onto the possible worlds based on a integer-valued function κ such that $\kappa(A \cup B) = \min(\kappa(A), \kappa(B))$. It is easy to check that $\Pi_\kappa(A) = 2^{-\kappa(A)}$ is a is a possibility degree. Moreover $\kappa(A) = n$ may be interpreted as a small probability of the form ϵ^n where ϵ is positive and close to 0.

4 Applications to Uncertainty Handling

This section describes some of the most common fields where possibility theory is used as a fruitful solution. It focuses in particular on statistics, where recent advances demonstrate the interest of a possibilistic approach.

4.1 Uncertainty Propagation

Uncertainty propagation considers a function $f : U_1 \times \ldots \times U_N \to U_{N+1}$ where the quantity $y = f(x_1, \ldots, x_n)$ presents some interest but there is uncertainty about $x_1, \ldots, x_n$. Possibilistic uncertainty propagation corresponds to consider a joint possibility distribution $\pi_{x_1,\cdots,x_n}$, possibly built from marginal ones on each x_i, and to propagate it through f to obtain a distribution π_y on the variable of interest.

The possibilistic approach based on Zadeh's extension principle applied to fuzzy intervals comes down to using level-wise interval propagation. As [Baudrit *et al.*, 2006] point out, it cannot be interpreted "as a conservative counterpart to the calculus of probabilistic variables under stochastic independence". One way out is to interpret fuzzy intervals as consonant belief functions and jointly propagate probabilistic and possibilistic information using both interval and Monte-Carlo computations, thus producing random intervals as an output.

Another approach recently presented by Hose and Hanss [Hose and Hanss, 2019b] proposes general aggregation operators combined with probability-possibility transformations (see Sec. 3.4) that are compatible with an interpretation of joint possibilities as imprecise probabilities, allowing one to also handle the case of dependent variables [Gray *et al.*, 2021] as well as the inverse problem [Hose and Hanss, 2019a], where the uncertainty has to be propagated from the output space to the input space. Such results suggest that one is not obliged to propagate probability and possibility each with their calculus, as done by [Baudrit *et al.*, 2006], thereby avoiding to produce fuzzy random variables (i.e., a probability distribution over possibility distributions), which remains hard to interpret for non-experts. Note that as f can be of any form, such propagation tools could be applied to check, e.g., the robustness of neural networks with respect to data perturbation [Hendrycks and Dietterich, 2019].

4.2 Statistical Inference and Machine Learning

One of the central goals of machine learning and statistical inference is to estimate the parameter θ of some function[1] $f_\theta : U_1 \times \ldots \times U_N \to U_{N+1}$, most commonly either by minimizing a loss function ℓ or maximizing the likelihood $P(d|\theta)$, whose strong connection with possibility theory is already mentioned in Sec. 3.2. In this problem as well, possibilistic approaches prove useful to treat incompleteness.

Uncertain data A first issue in which possibility distributions are instrumental is the modeling of uncertain data. Indeed, the lack of information concerning a observed piece of data (imprecision/incompleteness) is well handled by possibility distributions. If a piece of data $z = (\mathbf{x}, y)$ is replaced by some possibility distribution π_z, then there is a need to adapt either the likelihood or the loss function to ill-known observations [Couso and Dubois, 2018]. Authors have considered robust, min-max approaches [Guillaume and Dubois, 2020], as well as more optimistic, min-min approaches [Hüllermeier, 2014]. To be more precise, if we observe a piece of imprecise data $(\mathbf{X}, Y)$ where $\mathbf{X}$ and Y are subsets of possible values, and if $\ell : U_{N+1} \times U_{N+1} \to \mathbb{R}$ is a loss function, then we only know that the loss values lie within the set

$$\{\ell(f_\theta(\mathbf{x}), y) : \mathbf{x} \in \mathbf{X}, y \in Y\}.$$

Min-max approaches then retain the maximal values in the set, thus being pessimistic regarding the possible loss values,

[1] We focus here on predictive problems, but many works we cite also applies to generic distribution estimation.

and the min-min approaches, akin to semi-supervised learning, would take an optimistic stance. The possibilistic extension of such approaches would consist of taking the pessimistic/optimistic view level-wise [Guillaume and Dubois, 2015], and it can be shown that such approaches include some commonly used loss functions in robust statistics as special cases, such as the Huber one [Hüllermeier, 2014].

Due to its greater mathematical convenience and closeness to semi-supervised learning, the min-min approach has been successfully applied to machine learning, both to robustify model learning against noisy data [Lienen and Hüllermeier, 2024] and to perform self-supervised learning with possibilistic pseudo-labels used as robust replacements of unknown labels [Lienen and Hüllermeier, 2021]. In each of these, the use of possibilistic rather than probabilistic labels (called "soft") improved the learned model accuracy.

Model estimation and training It is fair to say that the crux of machine learning and statistical inference lies in parameter or model estimation, a task either achieved through (regularised) loss minimization or (Bayesian) probabilistic estimates. Here again, the possibilistic approach presents interesting methods, especially when dealing with limited samples, or with robust estimators.

In the case of loss functions, Serrurier and Prade [2013; 2015] have shown that one can adopt the imprecise probabilistic view of possibility theory given in Sec. 3.3 and leverage probability-possibility transforms of Sec. 3.4 to provide a possibilistic loss function consistent with a probabilistic approach, and providing improved results when data are scarce. More recently, Strauss et al. [2025] proposed a neuronal aggregation function based on maxitive measures, that can be readily embedded within classical deep learning and avoids the need for renormalization. Even if the authors admit that its potential remains to be explored, this example shows that possibilistic-based approaches can be made compatible with very standard machine learning approaches while not increasing computational costs.

In the case of likelihood functions and of a postulated true distribution $P(\cdot|\theta)$ generating the data, an essential question in both machine learning and statistics is how to obtain trustworthy, well calibrated estimates of the true value $\theta^* \in \Theta$ of the parameter. If the observed data $d = \{z^1, \dots, z^n\} \in \mathcal{D}$ are generated according to some process, one would like to derive, from d, confidence intervals or regions $C^d_\alpha \subseteq \Theta$ with strong statistical calibration properties, guaranteeing for instance that

$$\mathbf{P}(\theta^* \in C^d_\alpha) \geq \alpha \tag{2}$$

or, equivalently, that $\mathbf{P}(\theta^* \not\in C_\alpha) \leq 1 - \alpha$, meaning that as our confidence α increases, there is less and less probability that it does not contain the true parameter, and this with a controlled, calibrated probability. The recent trend of work around inferential models, led by Ryan Martin [2025], shows that such guarantees cannot (through an impossibility result) be obtained by an additive measure, but that maxitive measures can deliver the aforementioned calibrated estimators. The framework uses the relative likelihood $R(d, \theta) = \lambda(\theta)/\lambda(\theta_{MLE})$ where θ_{MLE} is the maximum likelihood estimate, and builds the possibility distribution

$$\pi^d(\theta) = P_\theta(\{D \in \mathcal{D} : R(D, \theta) \leq R(d, \theta)\})$$

which is normalized, and such that the sets $C^d_\alpha = \{\theta : \pi^d(\theta) \geq 1 - \alpha\}$ satisfy Equation (2). This approach is also connected to the recently introduced notion of E-value [Martin, 2023] that is currently gaining attraction within statistics.

In this case, not only possibility theory presents a valid choice among other options such as probability, evidence theory or imprecise probability, but it represents the most adequate solution to a problem that is unsolvable if one sticks to additive measures and probabilities.

Prediction Another focus of trustworthy artificial intelligence and machine learning is to provide uncertainty quantification of predictions. More precisely, given a predictive model f_θ and an observed test instance $\mathbf{x}_{test}$, can we provide tools in order not only to predict $f_\theta(\mathbf{x}_{test}) = y_{test}$, but also to equip the predictive method with an uncertainty quantification such that one predicts a subset $\mathcal{C} \subseteq U_{N+1}$ that provides some coverage guarantee of the ground-truth?

This is the main goal of the conformal calibration methods [Angelopoulos and Bates, 2023] that have now gained widespread attention within machine learning when it comes to uncertainty quantification, due to their combination of theoretical properties and easy-to-apply methodology. Conformal prediction relies on a number of exchangeable observations $z^1, \dots, z^c$ that are used to post-calibrate a given model f_θ in order to obtain prediction sets $\mathcal{C}_\alpha \subseteq U_{N+1}$ that benefit from the following coverage property [Angelopoulos and Bates, 2023, Appendix D]

$$P(Y_{test} \in \mathcal{C}_\alpha(X_{test})) \geq \alpha$$

where α is some prescribed probabilistic guarantee. While one can readily sees that the above inequality is remindful of Equation (1), Chella and Martin [2022] formally show that predictions derived by the conformal framework can be seen and interpreted as possibility distributions.

We think it quite interesting that the most recent developments within statistics tend to provide an important place to the very notion of possibility distribution and of possibility measure, suggesting that it has a key role to play in this area, without questioning the notion of probabilistic model.

4.3 Signal Processing

Signal processing is an important part of many areas of AI such as robotics, and possibility theory can in such cases be used to provide robust treatment of uncertainties.

Tracking and state estimation Tracking and state estimation are other very important problems within AI, with the celebrated Kalman filter coming first to mind. In this field as well, possibility theory was recently shown to provide an interesting alternative to probability in order to cope with incompleteness or with model misspecification, offering robust inference approaches.

In particular, [Houssineau and Bishop, 2018] recently developed a complete framework to perform temporal signal processing centered around the use of outer-probabilistic measures and the ∞-norm. They come down to using possibility measures and possibilistic calculus, providing counterparts to Markov process, Kalman filtering, signal smoothing, ... While the proposed framework can theoretically accommodate any continuous possibility distribution, a possibilistic counterpart of the Gaussian distribution is proposed:

$$\pi_{\bar{N}}(x; m, P) = \exp\left(-\frac{1}{2}(x-m)^T P^{-1}(x-m)\right)$$

with m the mode of the distribution and P a semi-definite positive matrix. This formulation is nothing else but a Gaussian distribution re-normalized so that its maximum be one. It is shown in [Houssineau and Bishop, 2018] that, combined with the proposed calculus based on the ∞-norm, one can obtain very efficient computational procedures.

In a subsequent series of paper, this framework was then shown to be quite effective in situations where information is lacking, or when robustness with respect to modeling assumptions was needed, such as when the field of view of sensors is limited in a decentralized, multi-sensor setting [Houssineau *et al.*, 2024] or when modeling assumptions are not met [Houssineau, 2021]. Of course, as probabilities and possibilities are at odds, this means that gaining on one aspect by switching the framework may imply a loss on another aspect. In the case of the well-known probability hypothesis density (PHD) filter, one loses the property of probabilistic expectation to have an efficient target counting tool, but one gains a higher robustness to misspecification. An interesting aspect to be explored would be to link this proposal with the possibilistic nets of Section 2.3, possibly extending them to general kinds of graphical models (it already applies to hidden Markov models that present a tree structure) as it would allow one to consider continuous possibilistic networks.

Robust signal filtering As recalled in Section 3.3, a possibility measure can be interpreted as a set of probabilities, from which can easily be extracted lower and upper expectation bounds. Convolution operators in image processing with specific kernels κ satisfy positivity ($\kappa > 0$) and normalisation ($\int \kappa = 1$); they are formally equivalent when computing expectation. Strauss et al. [2021] propose to view possibility distributions as sets of filters. Then, lower and upper expectation operators can be used to obtain lower and upper bounds of a filtered signal for which the corresponding filtering kernel is ill-known, e.g., has an ill-defined optimal bandwidth. Strauss et al. [2022] extend this approach to kernels satisfying normalization constraints but not necessarily positivity, through the use of two maxitive measures, thus allowing one to encompass well-known kernels that require negative values, such as those used to compute derivatives.

4.4 Preference Modeling

Preference modeling or learning is roughly divided into two different problems: either one wants to understand and model the observed preferences of a single user, to whom we can possibly ask tailored questions whose structure can vary from one question to the other, or one wants to understand the preferences of a given population of users in order to provide a generic recommendation, in which case one must learn a statistical model via structured and systematic observations that are easy to collect. Although it is defined on a structured space, the latter problem concerns statistical learning (see Section 4.2). In the former problem, uncertainty is necessarily due to incompleteness and imprecision, at least if one assumes that user preferences are stable and can be represented by a single model f_{θ^*}. In this case, uncertainty expressed by the user is purely subjective and not statistical, hence prone to possibilistic modelling. With this idea in mind, Adam and Destercke [2021] use possibility distributions to model uncertain user preferences in an incremental elicitation setting, thereby solving issues encountered by the classical set-valued approach.

Possibilistic approaches have also been used to model degrees of preferences themselves, not uncertainty about those. Eichhorn *et al.* [2016] proved that so-called Ordinal Conditional Functions networks (OCF-nets) can refine Conditional Preference (CP) - net orderings. Namely, they always lead to total orderings that are compatible with CP- nets. OCF-nets are based on ranking functions [Spohn, 2012] and obey an additive chain rule. As said in subsection 3.6, a ranking function can be easily turned into a possibility distribution, and product-based possibilistic networks, as OCF-nets can be also used for representing preferences; they are more flexible than CP-nets as they can also express conditional indifference [Ben Amor *et al.*, 2018].

5 Conclusion

An intelligent machine must be able to represent and reason under partial ignorance, and to know when it does not know. However large the data sets are, they cannot fully ban partial ignorance. Probability distributions alone cannot model ignorance. One needs to move to imprecise probability to capture it. However imprecise probability representations can be very complex. Possibility theory offers a simple representation of partial ignorance, despite its limited expressiveness.

This paper has recalled the roots and main elements of quantitative possibility theory, together with its link to the other main uncertainty theories. We have also reviewed the main current developments showing that quantitative possibility theory is at play in many different settings, including statistics, and offer a computationally attractive tool, in support to probabilistic methods rather than in competition with them, when one needs uncertainty reasoning that is robust to incompleteness. In these days of trustworthy AI, it can be an advantage in many applications, begging for further developments of possibilistic methods.

References

[Adam and Destercke, 2021] L. Adam and S. Destercke. Possibilistic preference elicitation by minimax regret. In *UAI*, volume 161 of *PMLR*, pages 718–727, 2021.

[Angelopoulos and Bates, 2023] A. Angelopoulos and S Bates. Conformal prediction:A gentle introduction. *Found. and Trends in Mach. Learn.*, 16(4):494–591, 2023.

[Ayachi *et al.*, 2014] R. Ayachi, N. Ben Amor, and S. Benferhat. A generic framework for a compilation-based inference in probabilistic and possibilistic networks. *Inf. Sci.*, 257:342–356, 2014.

[Baudrit *et al.*, 2006] C. Baudrit, D. Dubois, and D. Guyonnet. Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Trans. Fuzzy Syst.*, 14(5):593–608, 2006.

[Ben Amor *et al.*, 2018] N. Ben Amor, D. Dubois, H. Gouider, and H. Prade. Possibilistic preference networks. *Inf. Sci.*, 460-461:401–415, 2018.

[Benferhat and Smaoui, 2011] S. Benferhat and S. Smaoui. Inferring interventions in product-based possibilistic causal networks. *Fuzzy Sets Syst.*, 169(1):26–50, 2011.

[Benferhat and Tabia, 2010] S. Benferhat and K. Tabia. Belief revision of product-based causal possibilistic networks. *CanadianAI, LNCS*.6085, Springer, 244-255, 2010.

[Benferhat and Tabia, 2014] S. Benferhat and K. Tabia. Reasoning with uncertain inputs in possibilistic networks. In *KR*, pages 538 – 547. AAAI Press, 2014.

[Benferhat and Titouna, 2009] S. Benferhat and F. Titouna. Fusion and normalization of quantitative possibilistic networks. *Appl. Intell.*, 31(2):135–160, 2009.

[Benferhat *et al.*, 2002] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Bipolar possibilistic representations. In *UAI*, pages 45–52, 2002.

[Benferhat *et al.*, 2010] S. Benferhat, D. Dubois, H. Prade, and M.-A. Williams. A framework for iterated belief revision using possibilistic counterparts to Jeffrey's rule. *Fundam. Informaticae*, 99(2):147–168, 2010.

[Benferhat *et al.*, 2014] S. Benferhat, S. Lagrue, and K.Tabia. Interval-based possibilistic networks. In *SUM*, volume 8720 of *LNCS*, pages 37–50. Springer, 2014.

[Boutouhami *et al.*, 2017] K. Boutouhami, S. Benferhat, F. Khellaf, and F. Nouioua. Uncertain lightweight ontologies in a product-based possibility theory framework. *IJAR*, 88:237–258, 2017.

[Cella and Martin, 2022] L. Cella and R. Martin. Validity, consonant plausibility measures, and conformal prediction. *IJAR*, 141:110–130, 2022.

[Couso and Dubois, 2018] I. Couso and D. Dubois. A general framework for maximizing likelihood under incomplete data. *Int. J. Approx. Reason.*, 93:238–260, 2018.

[Denoeux, 2008] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artif. Intel.*, 172:234–264, 2008.

[Denoeux, 2014] Th. Denoeux. Likelihood-based belief function: Justification and some extensions to low-quality data. *IJAR*, 55:1535–1547, 2014.

[Destercke *et al.*, 2008] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations - I: Generalized p-boxes; II: Clouds. *IJAR*, 49:649–677, 2008.

[Dubois and Hüller-meier, 2007] D. Dubois and E. Hüller-meier. Comparing probability measures using possibility theory. *IJAR*, 45:364–385, 2007.

[Dubois and Prade, 1986] D. Dubois and H. Prade. A set-theoretic view of belief functions - Logical operations and approximation by fuzzy sets. *Int. J. of General Syst.*, 12:193–226, 1986.

[Dubois and Prade, 1988] D.Dubois and H.Prade.*Possibility Theory*.Plenum,1988.

[Dubois and Prade, 1992] D. Dubois and H. Prade.When upper probabilities are possibility measures.*Fuzzy Sets Syst.*, 49:65–74, 1992.

[Dubois and Prade, 1997] D. Dubois and H. Prade. Bayesian conditioning in possibility theory*Fuzzy Sets Syst.*, 92:223–240, 1997.

[Dubois and Prade, 1998] D. Dubois and H. Prade. Possibility theory: Qualitative and quantitative aspects. In D. Gabbay and P. Smets, editors, *Quantified Representation of Uncertainty and Imprecision*. 169-226, Kluwer, 1998.

[Dubois and Prade, 2016] D. Dubois and H. Prade. Practical methods for constructing possibility distributions. *Int. J. Intell. Syst.*, 31(3):215–239, 2016.

[Dubois and Prade, 2025] D. Dubois and H. Prade. 40 years of research in possibilistic logic - a survey. In *IJCAI*, pages 10427–10435. ijcai.org, 2025.

[Dubois *et al.*, 1993] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. *Fuzzy Logic: State of the Art*, pages 103–112, Kluwer, 1993.

[Dubois *et al.*, 1996] D. Dubois, H. Prade, and P. Smets. Representing partial ignorance. *IEEE Trans. Syst. Man Cybern. Part A*, 26(3):361–377, 1996.

[Dubois *et al.*, 1997] D. Dubois, S. Moral, and H. Prade. A semantics for possibility theory based on likelihoods. *J. Math. Anal. Appl.*, 205:359–380, 1997.

[Dubois *et al.*, 2004] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliab. Comput.*, 10:273–297, 2004.

[Dubois *et al.*, 2008] D. Dubois, H. Prade, and Ph. Smets. A definition of subjective possibility.*IJAR*,48:352–364,2008.

[Dubois *et al.*, 2016] D. Dubois, W. Liu, J. Ma, and H. Prade. The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. *Inf. Fusion*, 32:12–39, 2016.

[Dubois *et al.*, 2017] D. Dubois, G. Fusco, H. Prade, and A. G. B. Tettamanzi. Uncertain logical gates in possibilistic networks. *IJAR*, 82:101–118, 2017.

[Dubois *et al.*, 2020] D. Dubois, F.Faux, and H.Prade. Prejudice in uncertain information merging *IJAR*,121:1–22,2020.

[Dubois, 2006] D. Dubois. Possibility theory and statistical reasoning. *Comput. Stat. Data Anal.*, 51(1):47–69, 2006.

[Edwards, 1972] W.Edwards *Likelihood* Cambridge U.P.,1972.

[Eichhorn *et al.*, 2016] C. Eichhorn, M. Fey, and G. Kern-Isberner. CP-and OCF-networks – a comparison. *Fuzzy sets and Sys.*, 298:109–127, 2016.

[Fonck, 1997] P. Fonck. A comparative study of possibilistic conditional independence and lack of interaction. *IJAR*, 16:149–171, 1997.

[Gray *et al.*, 2021] A. Gray, D. Hose, M. de Angelis, M. Hanss, and S. Ferson. Dependent possibilistic arithmetic using copulas. *ISIPTA,PMLR*, 147:169–179, 2021.

[Guillaume and Dubois, 2015] R. Guillaume and D. Dubois. Robust parameter estimation of density functions under fuzzy interval observations. In *Proc. 9th ISIPTA*, pages 147–156, 2015.

[Guillaume and Dubois, 2020] R. Guillaume and D. Dubois. A min-max regret approach to maximum likelihood inference under incomplete data. *IJAR*, 121:135–149, 2020.

[Hendrycks and Dietterich, 2019] D. Hendrycks and T. Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *ICLR*, 2019.

[Higashi and Klir, 1982] M. Higashi and G. J. Klir. Measures of uncertainty and information based on possibility distributions. *Int. J. General Syst.*, 8:43–58, 1982.

[Hisdal, 1978] E. Hisdal. Conditional possibilities, independence and noninteraction*Fuzzy Sets Syst.*,1:283–297,1978.

[Hose and Hanss, 2019a] D. Hose and M. Hanss. Consistent inverse probability and possibility propagation. In *Proc. 11th EUSFLAT Conf.* , Prague, Atlantis Press, 2019.

[Hose and Hanss, 2019b] D. Hose and M. Hanss. Possibilistic calculus as a conservative counterpart to probabilistic calculus. *Mech. Syst. Signal Proc.*, 133:106290, 2019.

[Houssineau and Bishop, 2018] J.Houssineau and A.Bishop. Smoothing and filtering with a class of outer measures. *SIAM/ASA J. Uncert. Quantif.*, 6(2):845–866, 2018.

[Houssineau *et al.*, 2024] J. Houssineau, C.b. Xue, H. Cai, M. Uney, and E. Delande. Decentralised multi-sensor target tracking with limited field of view via possibility theory. In *FUSION*, pages 1–8. IEEE, 2024.

[Houssineau, 2021] J. Houssineau. A linear algorithm for multi-target tracking in the context of possibility theory. *IEEE Trans. Signal Process.*, 69:2740–2751, 2021.

[Hüllermeier, 2014] E. Hüllermeier. Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization*IJAR*, 55:1519–1534, 2014.

[Khellaf-Haned and Benferhat, 2014] H. F. Khellaf-Haned and S. Benferhat. Quantitative possibility theory: logical- and graphical-based representations. *J. Appl. Non Class. Logics*, 24(3):236–261, 2014.

[Lapointe and Bobée, 2000] S. Lapointe and B. Bobée. Revision of possibility distributions: A Bayesian inference pattern. *Fuzzy Sets Syst.*, 116:119–40, 2000.

[Levi, 1967] I. Levi. *Gambling with Truth. An essay in Induction and the Aims of Science*. MIT Press, 1967.

[Lienen and Hüllermeier, 2021] J.Lienen and E. Hüllermeier. Credal self-supervised learning*NeurIPS*,14370-14382.2021.

[Lienen and Hüllermeier, 2024] J.Lienen and E.Hüllermeier. Mitigating label noise through data ambiguation. In *AAAI*, pages 13799–13807, 2024.

[Martin, 2023] R. Martin. Ryan Martin's contribution to the discu-ssion of 'estimating means of bounded random variables by betting' by Waudby-Smith and Ramdas. *J.R.Stat. Soc.,Ser.B, Stat. Methodol.*, 86(1):43–44, 2023.

[Martin, 2025] R. Martin. Possibilistic inferential models: a review. arXiv:2507.09007 [math.st], 2025.

[Mauris, 2011] G. Mauris. Possibility distributions: A unified repre-sentation of usual direct-probability-based parameter estimation methods. *IJAR*, 52(9):1232–1242, 2011.

[Ramer, 1990] A. Ramer. Structure of possibilistic information metrics and distances:Properties & Possibilistic information metrics and distances:Characterizations of structure. *Int. J. Gen. Syst.*, 17 & 18:21–32 & 1–10, 1990.

[Serrurier and Prade, 2013] M. Serrurier and H. Prade. An informational distance for estimating the faithfulness of a possibility distribution, viewed as a family of probability distributions, with respect to data. *IJAR*, 54:919–933, 2013.

[Serrurier and Prade, 2015] M. Serrurier and H. Prade. Entropy evaluation based on confidence intervals of frequency estimates. In *ICML*, pages 1576–1584, 2015.

[Shackle, 1949] G. L. S. Shackle. *Expectation in Economics*. Cambridge University Press, 1949.

[Shafer, 1976] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[Shafer, 1987] G. Shafer. Belief functions and possibility measures. J. Bezdek, ed. In *Analysis of Fuzzy Information. 1: Mathematics and Logic*. 51-84, CRC Press, 1987.

[Smets, 1982] P. Smets. Possibilistic inference from statistical data. In *2nd World Conf. on Math. at the Service of Man*, pages 611–613. Univ. Politec. Las Palmas, 1982.

[Spohn, 2012] W. Spohn. *The Laws of Belief: Ranking Theory and Its Philosophical Applications*. Oxford U.P., 2012.

[Strauss *et al.*, 2021] O. Strauss, K. Loquin, and F. Kucharczak. On maxitive image processing. volume 394 of *Stud. Fuzziness Soft Comput.*, pages 201–215. Springer, 2021.

[Strauss *et al.*, 2022] O. Strauss, A. Rico, and Y. Hmidy. Macsum: A new in-terval-valued linear operator. *IJAR*, 145121–138, 2022.

[Strauss *et al.*, 2025] O. Strauss, A. Rico, J. Pasquet, and L. Pibre. Combining thresholded real values for designing an artificial neuron in a neural network. *Fuzzy Sets Syst.*, 499:109191, 2025.

[Walley, 1991] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[Walley, 1996] P. Walley. Measures of uncertainty in expert systems. *Artif. Intell.*, 83(1):1–58, 1996.

[Yager, 1983] R. R. Yager. An introduction to applications of possibility theory. *Human Syst. Mgmt.*, 3:246–269, 1983.

[Zadeh, 1978] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Syst.*, 1(1):3–28, 1978.

[Zadeh, 1979] L. A. Zadeh. A theory of approximate reasoning. In *Machine intelligence, V. 9*, pages 149–194. 1979.