Possibility Theory and Statistical Reasoning

Didier Dubois
Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier, 118 route de Narbonne
31062 Toulouse Cedex 4, France

May 15, 2006

Abstract

Numerical possibility distributions can encode special convex families of probability measures. The connection between possibility theory and probability theory is potentially fruitful in the scope of statistical reasoning when uncertainty due to variability of observations should be distinguished from uncertainty due to incomplete information. This paper proposes an overview of numerical possibility theory. Its aim is to show that some notions in statistics are naturally interpreted in the language of this theory. First, probabilistic inequalites (like Chebychev's) offer a natural setting for devising possibility distributions from poor probabilistic information. Moreover, likelihood functions obey the laws of possibility theory when no prior probability is available. Possibility distributions also generalize the notion of confidence or prediction intervals, shedding some light on the role of the mode of asymmetric probability densities in the derivation of maximally informative interval substitutes of probabilistic information. Finally, the simulation of fuzzy sets comes down to selecting a probabilistic representation of a possibility distribution, which coincides with the Shapley value of the corresponding consonant capacity. This selection process is in agreement with Laplace indifference principle and is closely connected with the mean interval of a fuzzy interval. It sheds light on the "defuzzification" process in fuzzy set theory and provides a natural definition of a subjective possibility distribution that sticks to the Bayesian framework of exchangeable bets. Potential applications to risk assessment are pointed out.

1 Introduction

There is a continuing debate in the philosophy of probability between subjectivist and objectivist views of uncertainty. Objectivists identify probabilities with limit frequencies and consider subjective belief as scientifically irrelevant. Conversely subjectivists consider that probability is tailored to the measurement of belief, and that subjective knowledge should be used in statistical inference. Both schools anyway agree on the fact that the only reasonable mathematical tool for uncertainty modelling is the probability measure. Yet, the idea, put forward by Bayesian subjectivists, that it is always possible to come up with a precise probability model, whatever the problem at hand looks debatable. This claim can be challenged due to the simple fact that there are at least two kinds of uncertain quantities: those which are subject to intrinsic variability (the height of adults in a country), and those which are totally deterministic but anyway ill-known, either because they pertain to the future (the date of the death of a living person), or just because of a lack of knowledge (I may not know in this moment the precise age of the President). It is clear that the latter cause of uncertainty is not "objective", because when lack of knowledge is at stake, it is always somebody's knowledge, which may differ from somebody else's knowledge. However, it is not clear that incomplete knowledge should be modelled by the same tool as variability (a unique probability distribution) [72, 42, 49]. One may argue with several prominent other scientists like Dempster [19] or Walley [108] that the lack of knowledge is precisely reflected by the situation where the probability of events is ill-known, except maybe for a lower and an upper bound. Moreover one may also have incomplete knowledge about the variability of a non-deterministic quantity if the observations made were poor, or if only expert knowledge is available. This point of view may to some extent reconcile subjectivists and objectivists: it agrees with subjectivists that human knowledge matters in uncertainty judgements, but it concedes to objectivists that such knowledge is generally not rich enough to allow for a full-fledged probabilistic modelling.

Possibility theory is one of the current uncertainty theories devoted to the handling of incomplete information, more precisely it is the simplest one, mathematically. To a large extent, it is similar to probability theory because it is based on set-functions. It differs from the latter by the use of a pair of dual set functions called possibility and necessity measures [28] instead of only one. Besides, it is not additive and makes sense on ordinal structures. The name "Theory of Possibility" was coined by Zadeh [119]. In Zadeh's view, possibility distributions were meant to provide a graded semantics to natural language statements. However, possibility and necessity measures can also be the basis of a full-fledged representation of partial belief that parallels probability. It can be seen either as a coarse, non-numerical version of probability theory, or a framework for reasoning with extreme probabilities (Spohn [102]), or yet a simple approach to reasoning with imprecise probabilities [36]. The theory of large deviations in probability theory also handles set-functions that look like possibility measures [90]. Formally, possibility theory refers to the study of *maxitive and minitive* set-functions, respectively called *possibility and necessity measures* such that the possibility degree of a disjunction of events is the maximum of the possibility degrees of events in the disjunction, and the necessity degree of a conjunction of events is the minimum of the necessity degrees of events in the conjunction. There are several branches of possibility theory, some being qualitative, others being quantitative, all satisfying the maxitivity and minitivity properties. But the variants of possibility theory differ for the conditioning operation. This survey focuses on numerical possibility theory. In this form, it looks of interest in the scope of coping with imperfect statistical information, especially non-Bayesian statistics relying on likelihood functions (Edwards, [46]), and confidence or prediction intervals. Numerical possibility theory provides a simple representation of special convex sets of probability functions in the sense of Walley [108], also a special case of Dempster's upper and lower probabilities [19], and belief functions of Shafer [96]. Despite its radical simplicity, this framework is general enough to model various kinds of information items: numbers, intervals, consonant (nested) random sets, as well as linguistic information, and uncertain formulae in logical settings [39].

Just like probabilities being interpreted in different ways (e.g., frequentist view vs. subjective view), possibility theory can support various interpretations. Hacking [63] pointed out that possibility can be understood either as an objective notion (referring to properties of the physical world) or as an epistemic one (referring to the state of knowledge of an agent). Basically there are four ideas each of which can be conveyed by the word 'possibility'. First is the idea of feasibility, such as ease of achievement, also referring to the solution to a problem, satisfying some constraints. At the linguistic level, this meaning is at work in expressions such as "it is possible to solve this problem". Another notion of possibility is that of plausibility, referring to the propensity of events to occur. At the grammatical level, this semantics is expressed by means of sentences such as "it is possible *that* the train arrives on time". Yet another view of possibility is logical and it refers to consistency with available information. Namely, stating that a proposition is possible means that it does not contradict this information. It is an all-or-nothing version of plausibility. The last semantics of possibility is deontic, whereby possible means allowed, permitted by the law. In this paper, we focus on the epistemic view of possibility, which also relies on the idea of logical consistency. In this view, possibility measures refer to the idea of plausibility, while the dual necessity functions attempt to quantify the idea of certainty. Plausibility is dually related to certainty, in the sense that the certainty of an event reflects a lack of plausibility of its opposite. This is a striking difference with probability which is self-dual. The expression *It is not probable that "not A"* is equivalent to saying *It is probable that A*, while the statement *It is not possible that "not A"* is not equivalent to saying *It is possible that A*. It has a stronger meaning, namely: *It is necessary that A*. Conversely, asserting that *it is possible that A* does not entail anything about the possibility nor the impossibility of "not A". Hence we need a dual pair of possibility and necessity functions.

There are not so many extensive works on possibility theory. The notion of epistemic possibility seems to appear during the 1940's in the work of the English economist G. L. S. Shackle [95], who called *degree of potential surprise* of an event its degree of impossibility, that is, the degree of certainty of the opposite event. The first book on possibility theory based on Zadeh's view [33] emphasises the close links between possibility theory and fuzzy sets, and mainly deals with numerical possibility and necessity measures. However it already points out some links with probability and evidence theories. Klir and Folger [76] insist on the fact that possibility theory is a special case of belief function theory, with again a numerical flavour. A more recent detailed survey [38] distinguishes between quantitative and qualitative sides of the theory. Basic mathematical aspects of qualitative possibility theory are studied at length by De Cooman [14]. More recently this author has investigated numerical possibility theory as a special case of imprecise subjective probability [16]. Qualitative possibility theory is also studied in detail by Halpern [64] in connection with other theories of uncertainty. The paper by Dubois, Nguyen and Prade [39] provides an overview of the links between fuzzy sets, probability and possibility theory. Uncertainty theories are also reviewed in the recent book by Klir [75] with emphasis on the study on information measures.

This paper is devoted to a survey of results in quantitative possibility theory, pointing out connections to probability theory and statistics. It however centers on the author's and close colleagues' views of the topic and does not develop mathematical aspects. The next section provides the basic equations of possibility theory. Section 3 shows how to interpret possibility distributions in the probabilistic setting. Section 4 bridges the gap between possibility distributions and confidence intervals. Section 5 envisages possibility theory from the standpoint of subjective probability and the principle of Insufficient Reason. Section 6 discusses the possibilistic counterpart to mathematical expectation. The last section points out the potential of possibility theory for uncertainty propagation in risk assessment.

2 Basics of Possibility Theory

The primitive object of possibility theory is the *possibility distribution*, which assigns to each element u in a set U of alternatives a degree of possibility $\pi(u) \in [0,1]$ of being the correct description of a state of affairs. This possibility distribution is a representation of what an agent knows about the value of some quantity x ranging on U (not necessarily a random quantity). Function π_x reflects the more or less plausible values of the unknown quantity x. These values are assumed to be mutually exclusive, since x takes on only one value (its true value). When $\pi_x(u) = 0$ for some u, it means that $x = u$ is considered an impossible situation. When $\pi_x(u) = 1$, it means that $x = u$ is just unsurprising, normal, usual, a much weaker statement than when probability is 1. Since one of the elements of U is the true value, the condition

$$\exists u, \pi_x(u) = 1$$

is assumed for at least one value $u \in U$. This is the normalisation condition. It claims that at least one value is viewed as totally possible. Indeed if $\forall u \in U, \pi_x(u) < 1$, the representation would be logically inconsistent because suggesting that all values in U are partially impossible for x. The degree of consistency of a submormalized possibility distribution is $cons(\pi) = \sup_{u \in U} \pi_x(u)$.

2.1 The Logical View

The simplest form of a possibility distribution on the set U is the characteristic function of a subset E of U, i.e.,$\pi_x(u) = 1$ if $x \in E, 0$ otherwise. It models the situation when all that is known about x is that it cannot lie outside E. This type of possibility distribution is naturally obtained from experts stating that a numerical quantity x lies between values a and b; then E is the interval $[a, b]$. This way of expressing knowledge is more natural than arbitrarily assigning a point-value u^* to x right away, because it allows for some imprecision.

Possibility as logical consistency has been put forward by Yager [113]. Consider a piece of incomplete information stating that "$x \in E$" is known for sure. This piece of information is incomplete insofar as E contains more than one element, i.e., the value of x can be any one (but only one) of the elements in E. Given such a piece of information, a set-function Π_E is built from E by the following procedure

$$\Pi_E(A) = \begin{cases} 1, & \text{if } A \cap E \neq \emptyset \ \ (x \in A \text{ and } x \in E \text{ are consistent}) \\ 0, & \text{otherwise } (A \text{ and } E \text{ are mutually exclusive}). \end{cases} \quad (1)$$

Clearly, $\Pi_E(A) = 1$ means that given that $x \in E$, $x \in A$ is possible because the intersection between set A and set E is not empty, while $\Pi_E(A) = 0$ means that $x \in A$ is impossible knowing that $x \in E$. It is easy to verify that such Boolean possibility functions Π_E satisfy the "maxitivity" axiom :

$$\Pi_E(A \cup B) = \max(\Pi_E(A), \Pi_E(B)). \quad (2)$$

Dually, a proposition is certain if and only if it logically derives from the available knowledge. Hence necessity also conveys a logical meaning in terms of deduction, just as possibility is a matter of logical consistency. The certainty of the event $x \in A$, knowing that $x \in E$, is then evaluated by the following index, called necessity measure:

$$N_E(A) = \begin{cases} 1, & \text{if } E \subseteq A \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Clearly the information $x \in E$ logically entails $x \in A$ when E is contained in A, so that certainty applies to events that are logically entailed by the available evidence. It can be easily seen that $N_E(A) = 1 - \Pi_E(A^c)$, denoting A^c the complement of A. In other words, A is necessarily true if and only if "not A" is impossible. Necessity measures satisfy the "minitivity" axiom

$$N_E(A \cap B) = \min(N_E(A), N_E(B)). \tag{4}$$

2.2 Graded possibility and necessity

However this binary representation is not entirely satisfactory. If the set E is too narrow, the piece of information is not so reliable. One is then tempted to use wide uninformative sets or intervals for the range of x. And sometimes, even the widest, safest set of possible values does not rule out some residual possibility that the value of x lies outside it. So it is natural to use a graded notion of possibility. Then, formally a possibility distribution π coincides with the membership function μ_F of a fuzzy subset F of U after Zadeh [117]. For instance, a fuzzy interval is a fuzzy set M of real numbers whose α-cuts $M_\alpha = \{u, \mu_M(u) \geq \alpha\}$ for $0 < \alpha \leq 1$, are nested intervals, usually closed ones (see Dubois, Kerre et al. [26] for an extensive survey). If u and u' are such that $\pi_x(u) > \pi_x(u'), u$ is considered to be a more plausible value than u'. The possibility degree of an event A, understood as a subset of U is then

$$\Pi(A) = \sup_{u \in A} \pi_x(u). \tag{5}$$

It is computed on the basis of the most plausible values of x in A, neglecting other realisations. The degree of necessity is then $N(A) = 1 - \Pi(A^c) = \inf_{u \notin A} 1 - \pi_x(u)$ [28, 33].

A possibility distribution π_x is at least as informative (we say *specific*) as another one π'_x if and only if $\pi_x \leq \pi'_x$ (see, e.g., Yager, [115]). In particular, if $\forall u \in \mathrm{U}, \pi_x(u) = 1, \pi_x$ contains no information at all (since it expresses that any value in U is possible for x). The corresponding possibility measure is then said to be *vacuous* and denoted $\Pi_\top$.

Remark. The possibility measure defined in (5) satisfies a strong form of maxitivity (2) for the union of infinite families of sets. On infinite sets, axiom (2) alone does not imply the existence of a possibility distribution satisfying (5). For instance, consider the natural integers, and a set function assigning possibility 1 to infinite subsets of integers, possibility 0 to finite subsets. This function is maxitive in the sense of (2) but does not fulfil (5) since $\pi(n) = \Pi(\{n\}) = 0, \forall n = 0, 1 \ldots$.

The possibilistic representation is capable of modelling several kinds of imprecise information within a unique setting. It is more satisfactory to describe imprecise numerical information by means of several intervals with various levels of confidence rather than a single interval. Possibility distributions can be obtained by extracting predictions intervals from probability measures as shown later on, or more simply by linear approximation between a core and a support provided by some expert. A possibility distribution π_x can more generally represent a finite family of nested confidence subsets $\{A_1, A_2, \ldots, A_m\}$ where $A_i \subset A_{i+1}, i = 1, \ldots, m-1$. Each confidence 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 $\lambda_1 \leq \ldots \leq \lambda_m$ due to the monotonicity of the necessity function N. The possibility distribution equivalent to the weighted family $\{(A_1, \lambda_1), (A_2, \lambda_2), \dots, (A_m, \lambda_m)\}$ is defined as the least informative possibility distribution π that obeys the constraints $\lambda_i = N(A_i), i = 1, \dots, m$. It comes down to maximizing the degrees of possibility $\pi(u)$ for all $u \in U$, subject to these constraints. The solution is unique and is, $\forall u$,

$$\pi_x(u) = \begin{cases} 1, & \text{if } u \in A_1 \\ \min_{i: u \notin A_i} 1 - \lambda_i, & \text{otherwise,} \end{cases} \tag{6}$$

which also reads

$$\pi_x(u) = \min_{i=1,\dots,m} \max(1 - \lambda_i, A_i(u)),$$

where $A_i(\cdot)$ is the characteristic function of A_i. The set of possibility values $\{\pi(u) : u \in U\}$ is then finite. This solution is the least committed one with respect to the available data, since by allowing the greatest possibility degrees in agreement with the constraints, it defines the least restrictive possibility distribution. Conversely, the family $\{(A_1, \lambda_1), (A_2, \lambda_2), \dots, (A_m, \lambda_m)\}$ of confidence intervals can be reconstructed from the possibility distribution π_x. Suppose that the set of possibility values $\pi_x(u)$ is $\{\alpha_1 = 1, \alpha_2 \geq \alpha_3 \dots \geq \alpha_m\}$ and let $\alpha_{m+1} = 0$. Then

$$A_i = \{u : \pi_x(u) \geq \alpha_i\}; \lambda_i = 1 - \alpha_{i+1}, \forall i = 1, \dots, m.$$

In particular, if $\lambda_m = 1$, then A_m is the subset which for sure contains x; moreover, $A_m = U$ if no strict subset of U surely includes x. This analysis extends to an infinite nested set of confidence intervals. Especially if M is a fuzzy interval with a continuous membership function, then the corresponding set of confidence intervals is $\{(M_\alpha, 1 - \alpha), 1 \geq \alpha > 0\}$. Denoting $M_\alpha(\cdot)$ the characteristic function of the cut M_α, it holds that $N(M_\alpha) = 1 - \alpha$, and $\mu_M(u) = \inf_{\alpha \in (0,1]} \max(\alpha, M_\alpha(u))(= \inf\{\alpha \in (0, 1] : \alpha > \mu_M(u)\})$.

2.3 Conditioning in possibility theory

Conditioning in possibility theory has been studied as a counterpart to probabilistic conditioning. However there is no longer a unique meaningful definition of conditioning, unlike in probability theory. Moreover the main difference between numerical and qualitative possibility theories lies in the conditioning process. The first notion of conditional possibility measure goes back to Hisdal [69]. She introduced the set function $\Pi(\cdot \mid A)$ through the equality

$$\forall B, B \cap A \neq \emptyset, \Pi(A \cap B) = \min(\Pi(B \mid A), \Pi(A)). \tag{7}$$

In order to overcome the existence of several solutions to this equation, the conditional possibility measure can be defined, as proposed by Dubois and Prade [33], as the least specific solution to this equation, that is, when $\Pi(A) > 0$ and $B \neq \emptyset$,

$$\Pi(B \mid A) = \left\{ \begin{array}{ll} 1, & \text{if } \Pi(A \cap B) = \Pi(A) \\ \Pi(A \cap B), & \text{otherwise .} \end{array} \right\} \tag{8}$$

The only difference with conditional probability is that the renormalisation via division is changed into a simple shift to 1 of the plausibility values of the most plausible elements in A. This form of conditioning agrees with a purely ordinal view of possibility theory and makes sense in a finite setting only. However, applied to infinite numerical settings, it creates discontinuities, and does not preserve the infinite maxitivity axiom. Especially $\Pi(B \mid A) = \sup_{u \in B} \pi(u \mid A)$ may fail to hold for non-compact events B [14]. The use of the product instead of minimum in the conditioning equation (7) enables infinite maxitivity to be preserved through conditioning. In close agreement with probability theory, it leads to

$$\Pi(B \mid A) = \frac{\Pi(A \cap B)}{\Pi(A)} \tag{9}$$

provided that $\Pi(A) \neq 0$. Then $N(B \mid A) = 1 - \Pi(B^c \mid A)$. See De Baets et al. [13] for a complete mathematical study of possibilistic conditioning, leading to the unicity of the product-based notion, in the infinite setting. The possibilistic counterpart to Bayes theorem looks formally the same as in probability theory:

$$\Pi(B \mid A) \cdot \Pi(A) = \Pi(A \mid B) \cdot \Pi(B). \tag{10}$$

However the actual expression of Bayes theorem in possibility theory is different, due to the maxitivity axiom and normalization. Consider the problem of testing hypothesis H against its complement, upon observing evidence E. One must compute $\Pi(H \mid E)$ in terms of $\Pi(E \mid H), \Pi(E \mid H^c), \Pi(H), \Pi(H^c)$ as follows:

$$\Pi(H \mid E) = \min(1, \frac{\Pi(E \mid H) \cdot \Pi(H)}{\Pi(E \mid H^c) \cdot \Pi(H^c)}) \tag{11}$$

where $\max(\Pi(H), \Pi(H^c)) = 1$. Moreover one must separately compute $\Pi(H^c \mid E)$. It is obvious that $\max(\Pi(H \mid E), \Pi(H^c \mid E)) = 1$. The merit of this counterpart to Bayes rule is that it does not require prior information to be available. In case of lack of such prior information, the uniform possibility $\Pi(H) = \Pi(H^c) = 1$ is used. Contrary to the probabilistic Bayesian approach, uninformative possibilistic priors are truly uninformative and invariant under any form of rescaling. It leads to compute

$$\Pi(H \mid E) = \min(1, \frac{\Pi(E \mid H)}{\Pi(E \mid H^c)}) \tag{12}$$

and $\Pi(H^c \mid E)$ likewise. It comes down to comparing $\Pi(E \mid H)$ and $\Pi(E \mid H^c)$, which corresponds to some existing practice in statistics, called likelihood ratio tests. Namely, the likelihood function is renormalized via a proportional rescaling [46]; [2]. This approach has been successfully developed for use in practical applications for instance by Lapointe and Bobée [81].

Yet another counterpart to Bayesian conditioning in quantitative possibility theory can be derived, noticing that in probability theory, $P(B \mid A)$ is an increasing function of both $P(B \cap A)$ and $P(A^c \cup B)$. This function is exactly $f(x, y) = \frac{x}{x+1-y}$. Then the following expression becomes natural ([37]))

$$\Pi(B \mid_f A) = \frac{\Pi(A \cap B)}{\Pi(A \cap B) + N(A \cap B^c)}. \tag{13}$$

The dual conditional necessity is such that

$$N(B \mid_f A) = 1 - \Pi(B^c \mid_f A) = \frac{N(A \cap B)}{N(A \cap B) + \Pi(A \cap B^c)}. \tag{14}$$

Interestingly, this conditioning also preserves consonance, and yields a possibility measure on the conditioning set A with distribution having support A:

$$\pi(u \mid_f A) = \begin{cases} \max(\pi(u), \frac{\pi(u)}{N(A)+\pi(u)}), & \text{if } u \in A \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

The $|_f$- conditioning clearly leads to a loss of specificity on the conditioning set A. In particular, if $N(A) = 0$ then π_A is totally uninformative on A. On the contrary, the product-based conditioning (9) operates a proportional rescaling of π on A, and corresponds to a form of information accumulation similar to Bayesian conditioning. It suggests that the product-based conditioning corresponds to a revision of the possibility distribution π into $\pi' = \Pi(\cdot \mid A)$, interpreting the conditioning event A in a strong way as $\Pi(A^c) = 0$ (A^c is impossible). On the other hand, the other conditioning $\Pi(B \mid_f A)$ is hypothetical, it evaluates the plausibility of B in the context where A is true without assuming that A^c is impossible. It corresponds to a form of contextual query-answering process based on the available uncertain information, focusing it on the reference class A. In particular, if π does not inform about A ($N(A) = 0$), then restricting to context A only produces uninformative results, that is $\pi(u \mid_f A)$ is vacuous on A while $\pi(u \mid A)$ is generally not so. A related topic is that of independence in possibility theory, which is omitted here for the sake of brevity. See Dubois et al. [39] for an extended discussion.

3 Relationship between probability and possibility theories

In quantitative theories on which we focus here, degrees of possibility are numbers that generally stand for upper probability bounds. Of course the probabilistic view is only one among other interpretive settings for possibility measures. Possibility degrees were introduced in terms of ease of attainment and flexible constraints by Zadeh [119] and as an epistemic uncertainty notion by Shackle [95], with little reference to probability and/or statistics. Levi [84, 85] was the first to relate Shackle's ideas to probability theory, in connection with Dempster's attempts to rationalize fiducial inferences [19]. Further on, Wang [112], Dubois and Prade [31] and others have developed a frequentist view of possibility, which suggests a bridge between possibility theory and statistical science. Possibility degrees then offer a simple approach to imprecise (set-valued) statistics, in terms of upper bounds of frequency. The comparison between possibility and probability theories is made easy by the parallelism of the constructs (i.e. the use of set-functions), which is not the case when comparing fuzzy sets and probability. As a mathematical object, maxitive set functions have been already studied by Shilkret, [99]. Possibility theory can also be viewed as a graded extension of modal logic where the dual notions of possibility and necessity already exist for a long time, in an all-or-nothing format. The notion "more possible than" was actually first modelled by David Lewis [86], in the setting of modal logics of counterfactuals, by means of a complete preordering relation among events satisfying some prescribed properties. This notion was independently rediscovered by Dubois [22] in the setting of decision theory, in an attempt to propose counterparts to comparative probability relations for possibility theory. Maxitive set-functions or equivalent notions have also emerged as a key tool in various domains, such as belief revision (Spohn [102]), non monotonic reasoning [6], game theory (the so-called unanimity games, Shapley [98]), imprecise probabilities (Walley [109]), etc. Due to the ordinal nature of the basic axioms of possibility and necessity functions, there is no enforced commitment to numerical possibility and necessity degrees. So there are basically two kinds of possibility theories: quantitative and qualitative [38]. This survey deals with quantitative possibility.

3.1 Imprecise Probability.

In the quantitative setting, we can interpret any pair of dual necessity/possibility functions $[N, \Pi]$ as upper and lower probabilities induced from specific convex sets of probability functions. Let π be a possibility distribution inducing a pair of functions $[N, \Pi]$. We define the probability family $\mathcal{P}(\pi) = \{P, \forall A \text{ measurable, } N(A) \leq P(A)\} = \{P, \forall A \text{ measurable, } P(A) \leq \Pi(A)\}$. In this case, $\sup_{P \in \mathcal{P}(\pi)} P(A) = \Pi(A)$ and $\inf_{P \in \mathcal{P}(\pi)} P(A) = N(A)$ hold (see [15, 36]). In other words, the family $\mathcal{P}(\pi)$ is entirely determined by the probability intervals it generates. Any probability measure $P \in \mathcal{P}(\pi)$ is said to be consistent with the possibility distribution π.

The pairs (interval A_i, necessity weight λ_i) suggested above 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 (like an interval containing the value of interest). These intervals can thus be obtained in terms of fractiles of a probability distribution. We define the corresponding probability family as follows: $\mathcal{P} = \{P, \forall A_i, \lambda_i \leq P(A_i)\}$. If the sets A_i are nested ($A_1 \subset A_2 \subset \ldots \subset A_n$, as can be expected for a family of confidence intervals), then $\mathcal{P}$ generates upper and lower probabilities of events that coincide with possibility and necessity measures induced by the possibility distribution (6) (see Dubois and Prade [36], De Cooman and Aeyels [15]) for details). Recently, Neumaier [89] has refined the possibilistic representation of imprecise probabilities using so-called 'clouds", which come down to considering the set of probability measures consistent with two possibility distributions. Walley and de Cooman [110] show that fuzzy sets representing linguistic information can be captured in the imprecise probability setting.

In a totally different context, well-known notions of probability theory such as probabilistic inequalities can be interpreted in the setting of possibility theory as well. Let x be the quantity to be estimated and suppose all is known is its mean value x^* and its standard deviation σ. Chebychev inequality defines bracketing approximations of symmetric intervals around x^* for unknown probability distributions, namely it can be proved that for any probability measure P having these characteristics,

$$P(X \in [x^* - a\sigma, x^* + a\sigma]) \geq 1 - \frac{1}{a^2} \text{ for } a \geq 1. \tag{16}$$

The preceding inequality suggests a simple method to build distribution-free possibilistic approximations of probability distributions [24], letting $\pi(x^* - a\sigma) = \pi(x^* + a\sigma) = \min(1, \frac{1}{a^2})$ for $a > 0$. It is easy to check that $P \in \mathcal{P}(\pi)$. Due to the nested structure of intervals involved in probabilistic inequalities, one may consider that such classical results can be couched in the terminology of possibility theory and yield convex sets of probability functions representable by possibility distributions. Of course such a representation based on Chebychev inequality is rather weakly informative.

Viewing possibility degrees as upper bounds of probabilities leads to the justification of the $|_f$-conditionalization of possibility measures. This view of conditioning can be called Bayesian possibilistic conditioning (Dubois and Prade [37]; Walley, [109]) in accordance with imprecise probabilities since $\Pi(B |_f A) = \sup\{P(B \mid A) : P(A) > 0, P \leq \Pi\}$. Such Bayesian possibilistic conditioning contrasts with product conditioning (9) which always supplies more specific results than the above. See De Cooman [16] for a detailed study of this form of conditioning and Walley and De Cooman [111] for yet other forms of conditioning in possibility theory.

3.2 Random Sets

From a mathematical point of view, the information modelled by π_x in (6), induced by the family of confidence sets $\{(A_1, \lambda_1), (A_2, \lambda_2), \ldots, (A_m, \lambda_m)\}$ can also be viewed as a nested random set. Indeed letting $\nu_i = \lambda_i - \lambda_{i-1}, \forall i = 1, \ldots, m+1$ (assuming the conventions: $\lambda_0 = 0; \lambda_{m+1} = 1; A_{m+1} = U$):

$$\forall u, \pi_x(u) = \sum_{i: u \in A_i} \nu_i. \tag{17}$$

The sum of weights $\nu_1, \ldots, \nu_{m+1}$ is 1. Hence the possibility distribution is the one-point coverage function of a random set [29], which can be viewed as a consonant belief function [97]. This view lays bare both partial ignorance (reflected by the size of the A_i's) and uncertainty (the ν_i's) contained in the information. And ν_i is the probability that the source supplies exactly A_i as a faithful representation of the available knowledge about x (it is not the probability that x belongs to A_i). Equation (17) still makes sense when the set of A_i's is not nested. This would be the case if the random set $\{(A_1, \nu_1), (A_2, \nu_2), \ldots, (A_{m+1}, \nu_{m+1})\}$ models a set of imprecise sensor observations along with the attached frequencies. In this case, and contrary to the nested situation, the probability weights ν_i can no longer be recovered from π using (17). However, this equation supplies an approximate representation of the data [34]. The random set view of possibility theory is developed in more details in (Joslyn [73], Gebhardt and Kruse, [55, 56, 57]). Note that the product-based conditioning of possibility measures formally coincides with Dempster rule of conditioning, specialised to possibility measures, i.e., consonant plausibility measures of Shafer [96]. It comes down to intersecting all sets A_i with the conditioning set A, assigning the weight ν_i to non-empty intersections $A_i \cap A$. Then, these weights are renormalized so that their sum be one.

A continuous possibility distribution on the real line can be defined by a probability measure on the unit interval (for instance the uniformly distributed one) and a multivalued mapping from $(0, 1]$ to $\mathcal{R}$, defining a family of nested intervals, following Dempster [19]. The probability measure is carried over from the unit interval to the real line via the multiple-valued mapping that assigns to each level $\alpha \in (0, 1]$ an α-cut of the fuzzy interval restricting an ill-known random quantity of interest (see Dubois and Prade [32], Heilpern [66], for instance). The possibility distribution is then viewed as a random interval I and $\pi(u) = P(u \in I)$ the probability that the value u belongs to a realization of I, as per equation (17).

3.3 Likelihood Functions

When prior probabilities are lacking, likelihood functions can be interpreted as possibility distributions, by default. Suppose a probabilistic model delivering the probability $P(u \mid y)$ of observing $x = u$, when a parameter $y \in V$, is fixed. Conversely, when $x = u^*$ is observed, the probability $P(u^* \mid v)$ is understood as the likelihood of $y = v$, in the sense that the greater $P(u^* \mid v)$ the more $y = v$ is plausible. It is easy to see that $\forall A \subseteq V$

$$\min_{v \in A} P(u^* \mid v) \leq P(u^* \mid A) \leq \max_{v \in A} P(u^* \mid v).$$

It is clear that the upper bound of the probability $P(u \mid A)$ is a possibility measure (see Dubois et al. [27]). Besides insofar as $P(u^* \mid A)$ is genuinely viewed as the likelihood of the proposition

$y \in A$, it is clear that whenever $A \subset B$, the proposition $y \in B$ should be at least as likely as the proposition $y \in A$. So by default it should be that $P(u^* \mid A) \geq \max_{v \in A} P(u^* \mid v)$ as well [11]. So, in the absence of prior information it is legitimate to define $\pi(v) = P(u^* \mid v)$, i.e. to interpret likelihood functions as possibility distributions. In the past, various people suggested to interpret fuzzy set membership degrees $\mu_F(v)$ as the probability $P('F' \mid v)$ of calling an element v an $'F'$ [70, 105].

The lower bound $\min_{v \in A} P(u^* \mid v)$ for $P(u^* \mid A)$ can be interpreted as a so-called "degree of guaranteed possibility" [43] $\Delta(A) = \min_{v \in A} \pi(v)$, which models the idea of evidential support of all realizations of event A. Viewing $P(u^* \mid v)$ as a degree of evidential support of v when observing u^*, $\Delta(A)$ is clearly the minimal degree of evidential support for $y \in A$, that evaluates to what extent all values in the set A are supported by evidence.

In general $\max_v P(u^* \mid v) \neq 1$ since in this approach, the normalisation with respect to v is not warranted. And in the continuous case $P(u^* \mid v)$ may be greater than 1. Yet, it is natural to assume that $\max_v P(u^* \mid v) = 1$. It corresponds to existing practice in statistics whereby the likelihood function is renormalized via a proportional rescaling (Edwards [46]). Besides, a value $y = v$ can be ruled out when $P(u^* \mid v) = 0$, while it is only highly plausible if $P(u^* \mid v)$ is maximal. If a value v of parameter y must be selected, it is legitimate to choose $y = v^*$ such that $P(u^* \mid v)$ is maximal. Hence possibilistic reasoning accounts for the principle of maximum likelihood. For instance, in the binomial experiment, observing $x = u^*$ in the form of k heads and $n - k$ tails for a coin where probability of heads is $y = v$, leads to consider the probability (= likelihood) $P(u^* \mid v) = v^k \cdot (1 - v)^{n-k}$, as the degree of possibility of $y = v$ and the choice $y = \frac{k}{n}$ for the optimal parameter value is based on maximizing this degree of possibility.

The product-based conditioning rule (9) can also be justified in the setting of imprecise probability, and maximum likelihood reasoning, noticing that, if event A, occurs one may interpret the probability degree $P(A)$ as the (second-order) degree of possibility of probability function $P \in \mathcal{P}(\pi)$, say $\pi'(P)$ in the face of the occurrence of A (the likelihood that P is the model governing the occurrence of A). Applying the maximum likelihood reasoning comes down to restricting $\mathcal{P}(\pi)$ to the set $\{P \leq \Pi, P(A) = \Pi(A)\}$ of probability functions maximizing the probability of occurrence of the observation A. It is clear that $\Pi(B \mid A) = \sup\{P(B \mid A), P(A) = \Pi(A) > 0, P \leq \Pi\}$ [42].

4 From confidence sets to possibility distributions

It is usual in statistics to summarize probabilistic information about the value of a model parameter θ by means of an interval containing this parameter with a degree of confidence attached to it, in other words a *confidence interval*. Such estimation methods assume the probability measure $P_x(\cdot \mid \theta)$ governing an observed quantity x to have a certain parameterized shape. The estimated parameter θ is expressed in terms of the observed data using a mathematical expression (an empirical estimator), and the distribution P_θ of the empirical evaluation of the parameter is then computed from the distribution of the measured quantity. Finally an interval $[a, b]$ such that $P_\theta(\theta \in [a, b]) \geq \alpha$ is extracted, where α is a prescribed confidence level. For instance, θ is the mean value of P_x, and P_θ is the distribution of the empirical mean $\theta^* = \frac{x_1 + \ldots + x_n}{n}$, based on n observations.

The practical application of this otherwise well-founded estimation method is often ad hoc. First, the choice of the confidence threshold is arbitrary (it is usually considered equal to 0.95, with no clear reasons). Second, the construction of the confidence interval often consists in locating $\frac{\alpha}{2}$- and $(1-\frac{\alpha}{2})$-fractiles, which sounds appropriate only if the distribution P_θ is symmetric. It is debatable because in the non-symmetric case (for instance using a χ^2 function for the estimation of variance) some excluded values may have higher density than other values lying in the obtained interval. Casting this problem in the setting of possibility theory may tackle both caveats. Namely, first, a possibility distribution can be constructed whose α-cuts are the confidence intervals of degree $1-\alpha$, and then such intervals can be located in a principled way, requesting that the possibility distribution be as specific as possible, thus preserving informativeness.

To simplify the framework, we consider the problem of determining *prediction intervals*, which consists in finding an interval containing, with a certain confidence level, the value of a variable x having a known distribution. The question is dual to the confidence interval problem, in the sense that in the latter the parameter θ is fixed, and the interval containing it changes with the chosen data sample, while, in the following, the interval is prescribed and the value it brackets is random.

4.1 Basic principles for probability-possibility transformations

Let P be a probability measure on a finite set U, obtained from statistical data. Let E be any subset of U. Define a possibility distribution π_E on U by letting

$$\pi_E(u) = \begin{cases} 1, & \text{if } u \in E \\ 1 - P(E), & \text{otherwise,} \end{cases} \tag{18}$$

It is obvious to check that by construction $\Pi_E(A) \geq P(A), \forall A \subseteq U$. Then the information represented by P can be approximated by the statement : "$x \in E$ with confidence at least $P(E)$", which is precisely encoded by π_E. Note that if the set E is chosen such that $P(E) = 0$, π_E is the vacuous possibility distribution. If E is too large, π_E is likewise not informative. In order for the statement "$x \in E$ with confidence $P(E)$" not to be trivial, $P(E)$ must be high enough, and E must be narrow enough. Since these two criteria are obviously antagonistic, one may either choose a confidence threshold α and minimise the cardinality of E such that $P(E) \geq \alpha$, or conversely fix the cardinality of E and maximize $P(E)$. Doing so for each value of confidence or each value of cardinality of the set is equivalent to finding a sequence $\{(E_1, \alpha_1), (E_2, \alpha_2), \ldots, (E_k, \alpha_k)\}$ of smallest prediction sets. If we let the probability measure be defined by $p(u_i) = p_i$ and assume $p_1 > \ldots > p_n > p_{n+1} = 0$, then it is obvious that the smallest set E with probability $P(E) \geq \sum_{j=1}^{i} p_j$ is $E = \{u_1, u_2, \ldots, u_i\}$. Similarly, $\{u_1, u_2, \ldots, u_i\}$ is also the set of cardinality i having maximal probability. So the set of most informative prediction sets is nested and is

$$\{(\{u_1\}, p_1), (\{u_1, u_2\}, p_1 + p_2), \ldots, (\{u_1, \ldots, u_i\}, \sum_{j=1}^{i} p_j) \ldots\}.$$

It comes down to turning a probability distribution P into a possibility distribution π^P of the following form ([29, 18]):

$$\pi_i^P = \sum_{j=i}^{n} p_j, \forall i = 1, \ldots, n, \tag{19}$$

denoting $\pi^P(u_i) = \pi_i^P$. This possibility distribution can be called the *optimal fuzzy prediction set* induced by P. Note that π^P is a kind of cumulative distribution function, with respect to the ordering on U defined by the probability values. Readers familiar with mathematical social sciences will notice it is a so-called Lorentz curve (see for instance Moulin[88]). It can be proved that this transformation of a probability measure into a possibility measure satisfies three basic principles (Dubois et al, [41]):

1. *Possibility-probability consistency*: one should select a possibility distribution π consistent with p, i.e., such that $P \in \mathcal{P}(\pi)$.
2. *Ordinal faithfulness* : the chosen possibility distribution should preserve the ordering of elementary events, namely, $\pi(u) > \pi(u')$ if and only if $p(u) > p(u')$. We cannot require the same condition for events since the possibility ordering is generally coarser than the probability ordering.
3. *Informativity*: The information content of π should be maximized so as to preserve as much from P as possible. It means finding the most specific possibility distribution in $\mathcal{P}(\pi)$. In the case of a statistically induced probability distribution, the rationale of preserving as much information as possible is natural.

It is clear that π^P satisfies all above three requirements and is the unique such possibility distribution. Note that the prediction sets organize around the mode of the distribution. The mode is indeed the most frequent value and is the most natural characteristics for an expert. When some elements in U are equiprobable the unicity of the possibility transform remains, but equation (19) must be applied to the well-ordered partition $U_1, \ldots, U_k$ of U induced by p, obtained by grouping equiprobable elements (elements in E_i have the same probability which is greater than the one of elements in E_{i+1}). Then equipossibility on each set U_i of the partition is assumed. For instance, the uniform probability is transformed into the uniform (vacuous) possibility distribution.

However it makes sense to relax the ordinal faithfulness condition into a weak form: $p(u) > p(u')$ implies $\pi(u) > \pi(u')$. Doing so, the most specific possibility transform is no longer unique. For instance, if $p_1 = \ldots = p_n = \frac{1}{n}$ then selecting any linear ordering of elements and applying (19) yields a most specific possibility distribution consistent with P. More generally if $U_1, \ldots, U_k$ is the well-ordered partition of U induced by p, then the most specific possibility distributions consistent with p are given by (19) applied to any linear ordering of U coherent with $U_1, \ldots, U_k$ (using an arbitrarily ranking of elements within each E_i).

The above approach has been recently extended by Masson and Denoeux [87] to the case when the empirical probability values p_i are parameters of a multinomial distribution, themselves estimated by means of confidence intervals. The problem is then to define a most specific possibility distribution covering all probability functions satisfying the constraints induced by the confidence intervals.

4.2 Alternative approaches to probability-possibility transforms

The idea that some consistency exists between possibilistic and probabilistic representations of uncertainty was suggested by Zadeh [119]. He defined the degree of consistency between a possibility distribution π and a probability measure P as follows: $Cons(P, \pi) = \sum_{i=1,n} \pi_i \cdot p_i$. It is the probability of the fuzzy event whose membership function is π. However Zadeh also described the consistency principle between possibility and probability in an informal way, whereby what is probable should be possible. Dubois and Prade [28] translated this requirement via the inequality $\Pi(A) \geq P(A)$ that founds the interpretation of possibility measures as upper probability bounds. There are two basic approaches to possibility/probability transformations. We presented one in the previous section 4.1, the other one is due to Klir [74, 58]. Klir's approach relies on a principle of information invariance, while the other one, described above, is based on optimizing information content [41]. Both respect a form of probability-possibility consistency. Klir tries to relate the notion of possibilistic specificity and the notion of probabilistic entropy. The *entropy* of a probability measure P is defined by

$$H(P) \;=\; -\sum_{j=1}^{n} p_j \cdot \log p_j. \tag{20}$$

In Klir's view, the transformation should be based on three assumptions :

1. A scaling assumption that forces each value π_i to be a function of p_i (where $p_1 \geq p_2 \geq \ldots \geq p_n$, that can be ratio-scale, interval scale, Log-interval scale transformations, etc.
2. An uncertainty invariance assumption according to which the entropy $H(p)$ should be numerically equal to some measure $E(\pi)$ of the information contained in the transform π of p.
3. Transformations should satisfy the consistency condition $\pi(u) \geq p(u), \forall u$.

The information measure $E(\pi)$ can be the logarithmic imprecision index of Higashi and Klir [68]

$$E(\pi) = \sum_{i=1}^{n} (\pi_i - \pi_{i+1}).Log_2 i, \tag{21}$$

or the measure of total uncertainty as the sum of two heterogeneous terms estimating imprecision and discord respectively (after Klir and Ramer [78]). The uncertainty invariance equation $E(\pi) = H(p)$, along with a scaling transformation assumption (e.g., $\pi(x) = \alpha p(x) + \beta, \forall x$), reduces the problem of computing π from p to that of solving an algebraic equation with one or two unknowns.

Klir's assumptions are debatable. First, the scaling assumption leads to assume that $\pi(u)$ is a function of $p(u)$ only. This pointwiseness assumption may conflict with the probability/possibility consistency principle that requires $\Pi \geq P$ for all events. See Dubois and Prade ([28], pp. 258-259) for an example of such a violation. Then, the nice link between possibility and probability, casting possibility measures in the setting of upper and lower probabilities cannot be maintained. The second and the most questionable prerequisite assumes that possibilistic and probabilistic information measures are commensurate. The basic idea is that the choice between possibility and probability is a mere matter of translation between languages "neither of which is weaker or stronger than the other" (quoting Klir and Parviz, [77]). It means that entropy and imprecision capture the same facet of uncertainty, albeit in different guises. The alternative approach recalled in section 4.1 does not make this assumption. Nevertheless Klir was to some extent right when claiming some similarity between entropy of probability measures and specificity of possibility distributions. In a recent paper [25], the following result is indeed established :

Theorem 1 *Suppose two probability measures P and Q such that π^P is strictly less specific than π^Q, then $H(P) > H(Q)$.*

It is easy to check that if π^P is strictly less specific than π^Q, then P is (informally) less peaked than Q (to use a terminology due to Birnbaum [7]), so that one would expect that the entropy of P should be higher than that of Q. Besides, viewing π^P as a kind of cumulative distribution, the comparison of probability measures via the specificity ordering of their possibility transforms is akin to a form of stochastic dominance called "majorization" in Hardy et al. [65]'s famous book on inequalities where a general result encompassing the above theorem is proved. This result lays bare a partial ordering between probability measures that seem to underlie many indices of dispersion (entropy, but Gini index as well and so on) as explained in [25], a result that actually dates back to the book of Hardy et al..

4.3 The continuous case

The problem of finding the best prediction interval for a random variable or best confidence intervals for a parameter does not seem to have received much attention, except for symmetric distributions. If the length L of the interval is prescribed, it is natural to maximise its probability. Nevertheless, it is not difficult to prove that, for a probability measure with a continuous unimodal density p, the optimal prediction interval of length L, i.e., the interval with maximal probability is $I_L = [a_L, a_L+L]$ where a_L is selected such that $p(a_L) = p(a_L + L)$. It is a cut $\{u, p(u) \geq \beta\}$ of the density, for a suitable value of a threshold β. This interval has degree of confidence $P(I_L)$ (often taken as 0.95) [41]. Conversely, the most informative prediction interval at a fixed level of confidence, say α is also of this form (choosing β such that $P(\{u, p(u) \geq \beta\}) = \alpha$. Clearly, in the limit, when the length L vanishes, I_L reduces to the mode of the distribution,viewed as the "most frequent value". It confirms that the mode of the distribution is the focus point of prediction intervals (not the mean), although such best prediction intervals are not symmetric around the mode. This result extends to multimodal density functions, but the resulting prediction set may consist of several disjoint intervals, since the best prediction set is of the form $\{u, p(u) \geq \beta\}$ for a suitable value of a threshold β[24]. It also extends to multidimensional universes [91].

Moving the threshold β from 0 to the height of the density, a probability measure having a unimodal probability density p with strictly monotonic sides can be transformed into a possibility distribution whose cuts are the prediction intervals of p with various confidence levels. The most specific possibility distribution π consistent with p, and ordinally equivalent to it, is obtained, such that [41]:

$$\forall L > 0, \pi(a_L) = \pi(a_L + L) = 1 - P(I_L).$$

Hence the α-cut of the optimal (most specific) π is the (1 - α)- prediction interval of p. These prediction intervals are nested around the mode of p. Going from objective probability to possibility thus means adopting a representation of uncertainty in terms of confidence intervals. Dubois, Mauris et al., [24] have found more results along this line for symmetric densities. Noticeably, each side of the optimal possibilistic transform is convex and there is no derivative at the mode of π because of the presence of a kink. Hence given a probability density on a bounded interval $[a, b]$, the symmetric triangular fuzzy number whose core is the mode of p and the support is $[a, b]$ is an upper approximation of p (in the style of, but much better than Chebychev inequality) regardless of its shape. In the case of a uniform distribution on $[a, b]$, any triangular fuzzy number with support $[a, b]$ provides a most specific upper approximation. These results and more recent ones (Baudrit et al. [3]) justify the use of triangular fuzzy numbers as fuzzy counterparts to uniform probability distributions, and model free approximations of probability functions with bounded support. This setting is also relevant for modeling sensor measurements [83].

A related problem is the selection of a modal value from a finite set of numerical observations $\{x_1 \leq \ldots \leq x_K\}$. Informally the modal value makes sense if a sufficiently high proportion of close observations exist. It is clear that such a modal value does not always exist, as in the case of the uniform probability. Dubois et al. [40] formalized this criterion as finding the interval length L for which the difference between the empirical probability value $P(I_L)$ for the most likely interval of length L, and the (uniform) probability $\frac{L}{L_{max}}$ is maximal, where $L_{max} = x_K - x_1$ is the length of the empirical support of the data. If L is very small or if it is very close to L_{max}, this difference is very small. The obtained interval achieves a trade-off between specificity (the length L of I_L) and frequency.

5 Possibility theory and subjective probability

So far, we have considered possibility measures in the scope of modeling information stemming from statistical evidence. There is a subjectivist side to possibility theory. Indeed, possibility distributions can arguably be advocated as a more natural representation of human uncertain knowledge than probability distributions. The representation of subjective uncertain evidence by a single probability function relies on interpreting probabilities as betting rates, in the setting of exchangeable bets. The probability $P(A)$ of an event is interpreted as the price of a lottery ticket an agent is ready to pay to a banker provided that this agent receives one money unit if event A occurs. A fair price is enforced, namely if the banker finds the buying price offered by the agent too small, they must exchange their roles. The additivity axiom is enforced using the Dutch book argument, namely the agent loses money each time the buying price proposal is such that $P(A) + P(A^c) \neq 1$. Then the uncertain knowledge of an agent must always be represented by a single probability distribution [82], however ignorant this agent may be.

However it is clear that degrees of probability obtained in the betting scheme will depend on the partition of alternatives among which to bet. In fact, two uniform probability distributions on two different frames of discernment representing the same problem may be incompatible with each other (Shafer [96]). Besides, if ignorance means not being able to tell if one contingent event is more or less probable than any other contingent event, then uniform probabilities cannot account for this postulate because, unless the frame of discernment is binary, even assuming a uniform probability, some contingent event will have a probability higher than another [42]. Worse, if an agent proposes a uniform probability as expressing his beliefs, say on facets of a die, it is not possible to know if this probability is the result of sheer ignorance, or if the agent really knows that the underlying process is random.

Several scholars, and noticeably Walley [108] challenged the unique probability view, and more precisely the exchangeable bet postulate. They admit the idea that an agent will not accept to buy a lottery ticket pertaining to the occurrence of A beyond a certain maximal price, nor to sell it under another higher price. The former buying price is interpreted as the lower probability $P_*(A)$, and the upper probability $P^*(A)$ is viewed as the latter selling price. Possibility measures can be interpreted in this way if their distribution is normalized. The possibilistic axiom $P_*(A \cap B) = \min(P_*(A), P_*(B))$ indicates an agent is not ready to put more money on A and on B than he would put on their conjunction. It indicates a cautious behavior. In particular, the agent would request free lottery tickets for all elementary events except at most one, namely the one with maximal possibility if unique (because $N(\{u\}) = 0, \forall u \in U$ whenever $\pi_1(u_1) = \pi_2(u_2) = 1$ for $u_1 \neq u_2$). This subjectivist view of possibility measures was first proposed by Robin Giles [60] in 1982, and developed by De Cooman and Aeyels [15].

The above subjectivist framework presupposes that decisions will be made on the basis of imprecise probabilistic information. Another view, whose main proponent is Smets [100], contends that one should distinguish between a credal level whereby an agent entertains beliefs and incomplete information is explicitly accounted for (in terms of belief functions according to Smets, but it could be imprecise probabilities likewise), and a so-called "pignistic" level where the agent should stick to classical decision theory when making decisions. Hence, this agent should use a unique probability distribution when computing expected utilities of such decisions. Under the latter view the question solved by Smets is how to derive a unique prior probability from a belief function. This approach is recalled here and related to Laplace Insufficient Reason principle. An interesting question is the converse problem : how to reconstruct the credal level from a prior probability distribution supplied by an expert.

5.1 A generalized Insufficient Reason principle

Laplace proposed that if elementary events are equally possible, they should be equally probable. This is the principle of Insufficient Reason that justifies, on behalf of respecting the symmetries of problems, the use of uniform probabilities over the set of possible states. Suppose there is a non-uniform possibility distribution on the set of states representing an agent knowledge. How to derive a probabilistic representation of this knowledge, where this dissymmetry would be reflected through the betting rates of the agent? Clearly, changing a possibility distribution into a probability distribution increases the informational content of the considered representation. Moreover, the probability distribution should respect the exiting symmetries in the possibility distribution, in agreement with Laplace principle.

A solution to this problem can be found by considering the more general case where the available information is represented by a random set $R = \{(A_1, \nu_1), (A_2, \nu_2), \ldots, (A_m, \nu_m)\}$. A generalised Laplacean indifference principle is then easily formulated: the weights p_i bearing on the focal sets are then uniformly distributed on the elements of these sets. So, each focal set A_i is changed into a uniform probability measure P^i over A_i, and the resulting probability PP is of the form $PP = \sum_{i=1}^{m} \nu_i \cdot P^i$. This transformation, already proposed by Dubois and Prade [29] comes down to a stochastic selection process that first selects a focal set according to the distribution ν, and then picks an element at random in the focal set A_i according to Laplace principle. The rationale behind this transformation is to minimize arbitrariness by preserving the symmetry properties of the representation.

Since a possibility distribution can be viewed as a nested random set, the pignistic transformation applies: the transformation of π yields a probability p_π defined by [30]:

$$p_\pi(\{u_i\}) = \sum_{j=i,\ldots,n} \frac{\pi_j - \pi_{j+1}}{j}, \tag{22}$$

where $\pi(\{u_i\}) = \pi_i$ and $\pi_1 \geq \ldots \geq \pi_n \geq \pi_{n+1} = 0$. This probability is also the gravity center of the set $\mathcal{P} = \{P \mid \forall A, P(A) \leq \Pi(A)\}$ of probability distributions dominated by Π [41]. Hence it can be viewed as applying the Insufficient Reason Principle to the set of probabilities $\mathcal{P}(\pi)$, equipping it with a uniformly distributed meta-probability, and then selecting the mean value.

This transformation coincides with the so-called pignistic transformation of belief functions (Smets [100]) as it really fits with the way a human would bet if possessing information in the form of a random set. Smets provides an axiomatic derivation of this pignistic probability: the basic assumptions are anonymity (permuting the elements of U should not affect the result) and linearity (the pignistic probability of a convex combination of random sets is the corresponding convex sum of pignistic probabilities derived from each random set). The pignistic probability also coincides with the Shapley value in game theory [98], where a cooperative game can be viewed as a non additive set function assigning a degree of strength to each coalition of agents. The obtained probability measure is then a depiction of the overall strength of each agent. Smets proposed similar axioms as Shapley.

5.2 A Bayesian approach to subjective possibility

If we stick to the Bayesian methodology of eliciting fair betting rates from the agent, but reject the credo 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 his beliefs can be more faithfully represented by a set of probabilities, the agent is forced to be additive by the postulates of exchangeable bets. Noticeably, the agent provides a uniform probability distribution whether (s)he knows nothing about the concerned phenomenon, or if (s)he knows the concerned phenomenon is purely random. In the Transferable Belief Model [101], the agent provides a pignistic probability induced by a belief function. Then, given a subjective probability, the problem consists in reconstructing the underlying belief function.

There are clearly several belief functions corresponding to a given pignistic probability. It is in agreement with intuition to consider the least informative among those. It means adopting a pessimistic view on the agent's knowledge. This is in contrast with the case of objective probability distributions where the available information is of statistical nature and should be preserved. Here, the available information being provided by an agent, it is not supposed to be as precise. One way of proceeding consists in comparing contour functions in terms of the specificity ordering of possibility distributions. Dubois et al. [44] proved that the least informative random set with a prescribed pignistic probability is unique and consonant. It is based on a possibility distribution π^{sub}, previously suggested in [30] with a totally different rationale:

$$\pi^{sub}(u_i) = \sum_{j=1,n} \min(p_j, p_i). \tag{23}$$

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 (17). 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 the pignistic transformation is a bijection between possibility and probability distributions. Equation (23) is also the transformation converse to eqn. (22). The subjective possibility distribution is less specific than the optimal fuzzy prediction interval (19), as expected, that is $\pi^{sub} > \pi_p$, generally. By construction, π^{sub} is a subjective possibility distribution. Its merit is not to assume human knowledge is precise, like in the subjective probability school. The transformation (23) was first proposed in [30] for objective probability, interpreting the empirical necessity of an event as summing the excess of probabilities of realizations of this event with respect to the probability of the most likely realization of the opposite event.

6 Fuzzy intervals and possibilistic expectations

A fuzzy interval [26] is a possibility distribution on the real line whose cuts are (generally closed) intervals. One interesting question is whether one can extract from such fuzzy intervals the kind of information useful in practice, that statisticians derive from probability distributions: cumulative distributions, mean values, variance, for instance. This section summarizes some known results on these issues.

6.1 Possibilistic cumulative distributions

Consider a fuzzy interval M with membership function μ_M viewed as a possibility distribution. The core of M is an interval $[m_*, m^*] = \{u, \mu_M(u) = 1\}$. The upper cumulative distribution function of the fuzzy interval M is $F^*(a) = \Pi_M((-\infty, a]) = \sup\{\mu_M(x) : x \leq a\}$, hence :

$$\forall a \in \mathcal{R}, F^*(a) = \begin{cases} \mu_M(a), & \text{if } a \leq m_* \\ 1, & \text{otherwise.} \end{cases} \tag{24}$$

Similarly, the lower distribution function $F_*(a) = N_M((-\infty, a]) = 1 - \Pi_M((a, +\infty)) = \inf\{1 - \mu_M(x) : x > a\}$ such that:

$$\forall a \in \mathcal{R}, F_*(a) = \begin{cases} 0, & \text{if } a < m^* \\ 1 - \lim_{x \to a_+} \mu_M(x), & \text{otherwise.} \end{cases} \tag{25}$$

The upper distribution function F^* matches the increasing part of the membership function of M. The lower distribution function F_* reflects the decreasing part of the membership function of M of which it is the fuzzy complement. In the imprecise probability view, M encodes a set of probability measures $\mathcal{P}(\mu_M)$. The upper and lower distribution functions are limits of distribution functions in this set [32]. However, the set of probability measures whose cumulative distributions lie between F_* and F^* is a superset of $\mathcal{P}(\mu_M)$ as indicated in [32] (see also [3]).

Ferson and Ginzburg [52] call a *p-box* a pair of cumulative distribution functions $(\underline{F}, \overline{F})$ with $\underline{F} \leq \overline{F}$. It can be viewed as a generalized interval as well. The above definitions shows that a fuzzy interval induces a p-box. But such generated p-boxes are less informative than the possibility distributions they are computed from. The point is that in the p-box view the two cumulative distributions are in some sense independent. They correspond to two random variables x^- and $x^+ > x^-$ defining the random interval $[x^-, x^+]$ with possibly independent end-points (See Heilpern [66, 67]; Gil [59]). Note that the intersection of all such generated intervals should not be empty so as to ensure the normalization of M. On the contrary, in the random set view, nested cuts $M_\lambda = [m_{*\lambda}, m^*_\lambda]$ are generated as a whole (hence a clear dependence between endpoints). Variables x^- and x^+ then depend on a single parameter λ such that $[x^-(\lambda), x^+(\lambda)] = [m_{*\lambda}, m^*_\lambda]$. In the p-box view, intervals of the form $[x^-(\alpha), x^+(\beta)]$ are generated for independent choices of α and β.

6.2 Possibilistic integrals

Possibility and necessity measures are very special cases of Choquet capacities that encode families of probabilities. The natural notion of integral in this framework is Choquet integral. A capacity is a monotone set-function σ (if $A \subseteq B$, then $\sigma(A) \leq \sigma(B)$), defined on an algebra $\mathcal{A}$ of subsets of U, with $\sigma(\emptyset) = 0$. The Choquet integral of a bounded function ϕ from a set U to the positive reals, with respect to a Choquet capacity (or fuzzy measure) σ, is defined as follows:

$$Ch_\sigma(\phi) = \int_0^\infty \sigma(\{u, \phi(u) \geq \alpha\})d\alpha, \tag{26}$$

provided that the cutset $\{u, \phi(u) \geq \alpha\}$ is in the algebra $\mathcal{A}$. See Denneberg [20] for a mathematical introduction. When $\sigma = P$, a probability measure, it reduces to a Lebesgue integral. When $\sigma = \Pi$, a possibility measure, and $\phi = \mu_F$ is the membership function of a fuzzy set F, also called a fuzzy event, it reads

$$Ch_\Pi(F) = \int_0^1 \Pi(F_\alpha)d\alpha = \int_0^1 \sup_u\{\mu_F(u) : \pi(u) \geq \alpha\}d\alpha. \tag{27}$$

In order to see that this identity holds, it suffices to notice (with De Cooman [16]) that the possibility degree of an event A can be expressed as $\Pi(A) = \int_0^1 \Pi_\top(A \cap \pi_\beta)d\beta$ where π_β is the β-cut of π and $\Pi_\top$ is the vacuous possibility measure ($\Pi_\top(A) = 1$ if $A \neq \emptyset$ and 0 otherwise). Then $Ch_\Pi(F)$ reads $\int_0^1 \int_0^1 \Pi_\top(F_\alpha \cap \pi_\beta)d\alpha d\beta$. It is allowed to commute the two integrals and the result follows, noticing that $\int_0^1 \Pi_\top(F_\alpha \cap \pi_\beta)d\beta = \sup\{\mu_F(u) : \pi(u) \geq \alpha\}$.

Equation (27) is a definition of the possibility of fuzzy events different from Zadeh's (namely $\Pi(F) = \sup_{u\in U} \min(\pi(u), \mu_F(u))$) because the maxitivity of $Ch_\Pi(F)$ w.r.t. F is not preserved here. That is, $Ch_\Pi(F \cup G) \neq \max(Ch_\Pi(F), Ch_\Pi(G))$ when $\mu_{F\cup G} = \max(\mu_F, \mu_G)$. The Choquet integral w.r.t. a necessity measure can be similarly defined, and it yields :

$$Ch_N(F) = \int_0^1 N(F_\alpha)d\alpha = \int_0^1 \inf_u\{\mu_F(u) : \pi(u) \geq \alpha\}d\alpha. \tag{28}$$

In order to see it, we use the duality between necessity and possibility, $N(F_\alpha) = 1 - \Pi((F_\alpha)^c)$ and denoting F^c the fuzzy complement of μ_F as $\mu_{F^c} = 1 - \mu_F$ we use the fact that $(F_\alpha)^c = \{u, \mu_{F^c}(u) > 1 - \alpha\}$. Plugging these identities into eqn. (27) yields the above expression of the Choquet integral of F w.r.t. a necessity measure. Moreover, the interval $[Ch_N(F), Ch_\Pi(F)]$ encloses all expectations of F w.r.t. all probabilities in $\mathcal{P}(\pi)$, boundaries being attained.

6.3 Expectations of fuzzy intervals

The simplest non-fuzzy substitute of the fuzzy interval M is its core, or its mode when its core is a singleton (in Dubois and Prade's early works [28], what is called mean value of a fuzzy number is actually its mode). Under the random set interpretation of a fuzzy interval, upper and lower mean values of M in the sense of Dempster [19], can be defined, i.e., $E_*(M)$ and $E^*(M)$, respectively, such that [32, 66]:

$$E_*(M) = \int_0^1 \inf M_\lambda d\lambda; \tag{29}$$

$$E^*(M) = \int_0^1 \sup M_\lambda d\lambda. \tag{30}$$

Note that these expressions are Choquet integrals of the identity function with respect to the possibility and the necessity measures induced by M. The mean interval of a fuzzy interval M is defined as $E(M) = [E_*(M), E^*(M)]$. It is thus the interval containing the mean values of all random variables compatible with M(i.e., $P \in \mathcal{P}(\mu_M)$). It is also the Aumann integral of the α-cut mapping : $\alpha \in (0, 1] \longrightarrow M_\alpha$, as recently proved by Ralescu [92].

That the mean value of a fuzzy interval be an interval seems to be intuitively satisfactory. Particularly the mean interval of a (regular) interval $[a, b]$ is this interval itself. The same mean interval obtains in the random set view and the imprecise probability view of fuzzy intervals, and is also the one we get by considering the cumulative distribution of the p-box induced by M. The upper and lower mean values are additive with respect to the fuzzy addition, since they satisfy, for u.s.c. fuzzy intervals [32, 66]:

$$E_*(M + N) = E_*(M) + E_*(N); \tag{31}$$

$$E^*(M + N) = E^*(M) + E^*(N), \tag{32}$$

where $\mu_{M+N}(z) = \sup_x \min(\mu_M(x), \mu_N(z - x))$. This property is a consequence of the additivity of Choquet integral for the sum of comonotonic functions.

6.4 The Mean Interval and Defuzzification.

Finding a scalar representative value of a fuzzy interval is often called defuzzification in the literature of fuzzy control (See Yager and Filev [116], and Van Leekwijk and Kerre [107] for extensive overviews). Various proposals exist:

- the mean of maxima (MOM), which is the middle point in the core of the fuzzy interval M,

- the center of gravity. This is the center of gravity of the support of M, weighted by the membership grade.
- the center of area (median): This is the point of the support of M that equally divides the area under the membership function.

The MOM sounds natural as a representative of a fuzzy interval M in the scope of possibility theory where values of highest possibility are considered as default plausible values. This is in the particular case when the maximum is unique. However, the MOM clearly does not exploit all the information contained in M since it neglects the membership function. Yager and Filev [116] present a general methodology for extracting characteristic values from fuzzy intervals. They show that all methods come down to a possibility-probability transformation followed by the extraction of characteristic value such as a mean value. Note that the MOM, the center of gravity and the center of area come down to renormalizing the fuzzy interval as a probability distribution and computing its mode, expected value or its median, respectively. These approaches are ad hoc. Moreover, the renormalization technique (dividing the membership function by its surface) is itself arbitrary since the obtained probability may fail to belong to $\mathcal{P}(\mu_M)$, the set of probabilities dominated by the possibility measure attached to M [28]. In view of the quantitative possibility setting, it seems that the most natural defuzzication proposal is the middle point of the mean interval [114]

$$\overline{E}(M) = \int_0^1 \frac{(\inf M_\lambda + \sup M_\lambda)}{2} d\lambda = \frac{E_*(M) + E^*(M)}{2}. \tag{33}$$

Only the mean interval accounts for the specific possibilistic nature of the fuzzy interval. The choice of the middle point expresses a neutral attitude of the user and extends the MOM to an average mean of cut midpoints. Other choices are possible, for instance using a weighted average of $E_*(M)$ and $E^*(M)$. Fullér and colleagues [9, 53] consider introducing a weighting function on $[0, 1]$ in order to account for unequal importance of cuts when computing upper and lower expectations.

$\overline{E}(M)$ has a natural interpretation in terms of simulation of a "fuzzy variable". Chanas and Nowakowski [10] investigate this problem in greater detail. Namely, consider the two step random generator which selects a cut at random (by choosing $\lambda \in (0, 1]$), and a number in the cut M_λ. The corresponding random quantity is $x(\alpha, \lambda) = \alpha \cdot \inf M_\lambda + (1 - \alpha) \cdot \sup M_\lambda$. The mean value of this random variable is $\overline{E}(M)$ and its distribution is P_M with density [41] :

$$p_M(x) = \int_0^1 \frac{M_\lambda(x)}{\sup M_\lambda - \inf M_\lambda} d\lambda. \tag{34}$$

The probability distribution P_M is in fact, the center of gravity of $\mathcal{P}(\mu_M)$. It corresponds to the pignistic transformation of M, obtained by considering cuts as uniformly distributed probabilities. The mean value $\overline{E}(M)$ is linear in the sense of fuzzy addition and scalar multiplication [94, 54].

6.5 The variance of a fuzzy interval

The notion of variance has been extended to fuzzy random variables [79], but little work exists on the variance of a fuzzy interval. Fuller and colleagues [9, 53] propose a definition as follows:

$$\overline{V}(M) = \int_0^1 (\frac{\sup M_\lambda - \inf M_\lambda}{2})^2 f(\lambda) d\lambda, \quad (35)$$

where f is a weight function. However appealing this definition may sound, it lacks proper interpretation in the setting of imprecise probability. In fact, the very idea of a variance of a possibility distribution is somewhat problematic. A possibility distribution expresses information incompleteness, and does not so much account for variability. The variance of a constant but ill-known quantity makes little sense. The amount of incompleteness is then well-reflected by the area under the possibility distribution, which is a natural characteristics of a fuzzy interval. Other indices of information already mentioned in section 4.2 are variants of this simpler index. Additionally, it is clear that the expression of $\overline{V}(M)$ depends upon the area under the possibility distribution (suffices to let $f(\lambda) = 1$). So it is not clear that the above definition qualifies as a variance: the wider a probability density, the higher the variability of the random variable, but the wider a fuzzy interval, the more imprecise.

However if the possibility distribution $\pi = \mu_M$ stands for subjective knowledge about an ill-known random quantity, then it is interesting to compute the range $V(M)$ of variances of probability functions consistent with M, namely

$$V(M) = \{\text{variance}(P), P \in \mathcal{P}(\mu_M)\}. \quad (36)$$

Viewing a fuzzy interval as a set of cuts, determining $V(M)$ is closely related to the variance of a finite set of interval-valued data, for which results and computational methods exist [50]. In the nested case, which applies to fuzzy intervals, the determination of the upper bound of the variance interval is a NP-hard problem, while the lower bound is trivially 0. The upper bound of $V(M)$ can be called the potential variance of M, and it is still an open problem to find a closed form expression for it in the general case, let alone its comparison with the definition (35) by Fuller. In the symmetric case, Dubois et al. [23] proved that the potential variance of M is precisely given by (35) when $f(\lambda) = 1$. We can conjecture that this result holds in the general case, which would suggest that the amount of imprecision in the knowledge of a random variable reflects its maximal potential range of variability. Also of interest is to relate the potential variance to the scalar variance of a fuzzy random variable, introduced by Koerner [79].

7 Uncertainty propagation with possibility distributions

A very important issue for applications of possibility theory is that of propagating uncertainty through mathematical models of processes. Traditionally this problem was addressed as one of computing the output distribution of functions of random variables using Monte-Carlo methods typically. Monte-Carlo methods applied to the calculation of functions of random variables cannot account for all types of uncertainty propagation [48]. The necessity of different tools for telling variability from partial ignorance leads to investigate the potential of possibility theory in addressing this issue, since this theory is tailored to the modeling of incomplete information. Suppose there are $k < n$ random variables $(X_1, \ldots, X_k)$ and $n - k$ possibilistic quantities $(X_{k+1}, \ldots, X_n)$. The problem, already outlined by Kaufmann and Gupta [62] is to compute the available information on a function $f(X_1, \ldots, X_n)$.

First assume that only two possibilistic quantities are present ($k = 0, n = 2$). The extension principle of Zadeh [118] proposes a solution to the propagation problem in the setting of possibility theory [33]. Consider a two place function f. If a joint possibility relating two ill-known quantities x_1 and x_2 is separable, i.e., $\pi = \min(\pi_1, \pi_2)$, then the possibility distribution π_f of $f(x_1, x_2)$ is

$$\pi_f(v) = \begin{cases} \sup\{\min(\pi_1(u_1), \pi_2(u_2)) : f(u_1, u_2) = v\}, & \text{if } f^{-1}(\{v\}) \neq \emptyset; \\ 0, & \text{otherwise .} \end{cases} \tag{37}$$

This proposal lays bare a basic issue that has to be solved prior to performing the propagation step: how to represent joint possibility distributions and what type of independence is involved? In the purely possibilistic framework, the choice $\pi = \min(\pi_1, \pi_2)$ for the joint distribution reflects a principle of minimal specificity: the largest, least committed joint possibility distribution whose projections are π_1 and π_2 is $\pi = \min(\pi_1, \pi_2)$. So the above calculation presupposes nothing on the possible dependence between quantities x and y. Moreover, the extension principle applied to non-fuzzy possibility distributions (valued on $\{0, 1\}$) reduces to standard interval calculations.

In the setting of random sets [57], the partial lack of information gives birth two possible levels of dependence (that can hardly be told apart using single probability distributions): on top of the possible dependence between variables, one must consider the possible dependence between the sources of information. The above notion of joint possibility distributions relies on a dependence assumption between sources, but no dependence assumption between variables is made. It presupposes that if the first source delivers a cut $(A_1)_\lambda$ of π_1 then the other one delivers $(A_2)_\lambda$ for the same value of λ. Two nested random sets with associated one-point coverage functions π_1 and π_2 then produce a nested random set with one-point coverage function $\min(\pi_1, \pi_2)$. In other words, it comes down to working with confidence intervals having the same levels of confidence. For instance, if π_1 and π_2 are supplied by the same expert, such a dependence assumption between confidence levels looks natural.

On the contrary, the assumption that quantities x_1 and x_2 are independently observed will not lead to a nested random set, since cuts $(A_1)_\lambda$ and $(A_1)_\rho$ are compatible, for $\lambda \neq \rho$. Hence the set of joint observations is not equivalent to a joint possibility distribution [35].

Let π_i be associated with the nested random interval with mass assignment function ν_i for $i = 1, 2$. Let $\nu(A, B)$ be the joint mass assignment whose projections are ν_i for $i = 1, 2$. That is:

$$\begin{aligned} \nu_1(A) = & \textstyle\sum_B \nu(A, B); \\ \nu_2(B) = & \textstyle\sum_A \nu(A, B). \end{aligned} \tag{38}$$

Note that since $\nu(A, B)$ is assigned to the Cartesian product $A \times B$, there is still no dependence assumption made between the variables x_1 and x_2. Assuming independence between the sources of information leads to define a joint random set describing (x_1, x_2) by means of Dempster rule of combination of belief functions, that is, $\nu(A, B) = \nu_1(A) \cdot \nu_2(B)$. The random set induced on $f(x_1, x_2)$ has mass function ν_f defined by

$$\nu_f(C) = \sum_{A,B: f(A,B)=C} \nu_1(A) \cdot \nu_2(B), \tag{39}$$

where $f(A, B) = C$ is obtained by interval analysis. The one-point coverage function of ν_f (equation 17) can be directly expressed in terms of π_1 and π_2 by the sup-product extension principle (changing minimum into product in (37) [34]. The random set setting for computing with possibility distributions thus encompasses both fuzzy interval and random variable computation. In practice the above calculation can be carried out on continuous possibility distributions using a Monte-Carlo method that selects α-cuts of π_1 and π_2, and interval analysis on selected cuts.

A total absence of knowledge about dependence between sources may also be assumed. If the joint mass function $\nu(A, B)$ is unknown, a more heavy computation scheme can be invoked. Namely, for any event C of interest it is possible to compute probability bounds induced by the only knowledge of the marginal distributions π_1 and π_2. For instance the lower probability bound can be obtained by minimizing $P_*(C) = \sum_{A,B:f(A,B)\subseteq C} \nu(A, B)$ under the constraints (38), and the upper probability bound by maximizing $P^*(C) = \sum_{A,B:f(A,B)\cap C\neq\emptyset} \nu(A, B)$ under the same constraints. The obtained bounds are the most conservative one may think of [4].

When k variables are probabilistic, other quantities being possibilistic, one may perform Monte-Carlo sampling of random variables $(X_1, \ldots, X_k)$ and fuzzy interval analysis on possibilistic quantities $(X_{k+1}, \ldots, X_n)$ [47, 61]. This presupposes that random variables and possibilistic quantities are independently informed, random variables being mutually independent and possibilistic variables depending on a single source of information(for instance several sensors and a human expert, respectively). Then $f(X_1, \ldots, X_n)$ is a fuzzy random variable for which average upper and lower cumulative distributions can be derived[5]. Such kinds of hybrid calculations were implemented by Ferson[47], Guyonnet et al. [61] in the framework of risk assessment for pollution studies, and geology [1]. However the above schemes making other kinds of assumptions between possibilistic variables can be accommodated in a straightforward way to the hybrid probability/possibility situation. The above outlined propagation methods should be articulated in a more precise way with other techniques which propagate imprecise probabilities in risk assessment models [51, 71]. Note that casting the above calculations in the setting of imprecise probabilities enables dependence assumptions to be made between marginal probabilities dominated by π_1 and π_2, while the random set approach basically accounts for (in)dependence assumptions between the sources of information. The study of independence when both variability and lack of knowledge are present is not yet fully understood (see for instance Couso et al. [12]).

If the user of such methods is interested by the risk of violating some threshold for the output value, upper and lower cumulative distributions can be derived. This mode of presentation of the results lays bare the distinction between lack of knowledge (the distance between the upper and lower cumulative distributions) and variability (the slopes of the cumulative distributions influenced by the variances of the probabilistic variables). These two dimensions would be mixed up in the variance of the output if all inputs were represented by single probability distributions. Besides note that upper and lower cumulative distributions only partially account for the actual uncertain output (contrary to the classical probability case). So more work is needed to propose simple representations of the results that may help the user exploit this information. The proper information to be extracted clearly depends on the question of interest (for instance determining the likelihood that the output value lies between two bounds cannot be addressed by means of the upper and lower cumulative distributions).

8 Conclusion

Quantitative possibility theory seems to be a promising framework for probabilistic reasoning under incomplete information. This is because some families of probability measures can be encoded by possibility distributions. The simplicity of possibility distributions make them attractive for practical applications of imprecise probabilities, and more generally for the representation of poor probabilistic information. Besides, cognitive studies for the empirical evaluation of possibility theory have recently appeared [93]. Their experiments suggest "that human experts might behave in a way that is closer to possibilistic predictions than probabilistic ones". The cognitive validation of possibility theory is clearly an important issue for a better understanding of when possibility theory is most appropriate.

The connection between possibility and probability in the setting of imprecise probabilities makes it possible to envisage a unified approach for uncertainty propagation with heterogeneous information, some sources providing regular statistics, other ones subjective information under the form of ill-defined probability densities [71]. Other applications of possibilistic representations of poor probabilistic knowledge are in measurement [83].

Moreover a reassessment of non-Bayesian statistical methods in the light of possibility theory seems to be promising. This paper only hinted to that direction, focusing on some important concepts in statistics, such as confidence intervals and the maximum likelihood principle. However much work remains to be done. In the case of descriptive statistics, a basic issue is the handling of set-valued or fuzzy data: how to extend existing techniques ranging from the calculation of empirical mean, variances and correlation indices [21], to more elaborated data analysis methods such as clustering, regression analysis [45], principal component analysis and the like. Besides, possibility theory can offer alternative representations of classical data sets, for instance possibilistic clustering [80, 106] where class-membership is graded, fuzzy regression analysis where a fuzzy-valued affine function is used [103, 104]. Concerning inferential statistics, the use of possibility theory suggests substituting a single probabilistic model with a concise representation of a set of probabilistic models among which the user is not forced to choose if information is lacking. The case of scarce empirical evidence is especially worth studying. For instance, Masson and Denoeux [87] consider multinomial distributions induced by a limited number of observations from which possibly large confidence intervals on probabilities of realizations are obtained. These intervals delimit a set of potential probabilistic models to be encompassed by means of a single possibility distribution. Since confidence intervals are selected using a threshold, doing away with this threshold leads to a higher order possibility distribution on probabilistic models, similar to De Cooman's approach to fuzzy probability intervals[17]. Lastly, the representation of high-dimensional possibility measures can be envisaged using the possibilistic counterpart to Bayesian networks [8].

Acknowledgements This paper is based on previous works carried out with several people, especially Henri Prade, Gilles Mauris, Philippe Smets, and more recent works with Cedric Baudrit and Dominique Guyonnet.

References

[1] Bardossy G., Fodor J., 2004. Evaluation of Uncertainties and Risks in Geology, Springer, Berlin.

[2] Barnett V., 1973. Comparative Statistical Inference, J. Wiley, New York.

[3] Baudrit C., Dubois D., Fargier H., 2004. Practical representation of incomplete probabilistic information. In: M. Lopez-Diaz et al. (Eds) Soft Methodology and Random Information Systems (Proc. 2nd International Conference on Soft Methods in Probability and Statistics, Oviedo), Springer, Berlin, 149-156. To appear in Comput. Stat. & Data Anal.

[4] Baudrit. C., Dubois D., 2005. Comparing Methods for Joint Objective and Subjective Uncertainty Propagation with an example in a risk assessment, Proc. 4th Int. Symposium on Imprecise Probabilities and Their Applications Pittsburgh, USA, 2005, 31-40.

[5] Baudrit. C., Guyonnet D., Dubois D., 2005. Post-processing the hybrid method for addressing uncertainty in risk assessments, J. Environmental Engineering, 131,1750-1754.

[6] Benferhat S., Dubois D. and Prade H., 1997. Nonmonotonic reasoning, conditional objects and possibility theory, Artificial Intelligence, 92, 259-276.

[7] Birnbaum Z. W., 1948. On random variables with comparable peakedness, Annals of Mathematical Statistics, 19, 76-81.

[8] Borgelt C., and Kruse R., 2002. Learning from imprecise data: possibilistic graphical models Computational Statistics & Data Analysis, 38, 449-463.

[9] Carlsson C., and Fuller R., 2001. On possibilistic mean value and variance of fuzzy numbers Fuzzy Sets and Systems, 122, 315-326.

[10] Chanas S. and Nowakowski M., 1988. Single value simulation of fuzzy variable, Fuzzy Sets and Systems, 25, 43-57.

[11] Coletti G. and Scozzafava R., 2003. Coherent conditional probability as a measure of uncertainty of the relevant conditioning events, Proc. ECSQARU03, Aalborg, LNAI vol. 2711, Springer Verlag, Berlin, 407-418.

[12] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities Risk Decision and Policy **5** 165–181, 2000.

[13] De Baets B., Tsiporkova E. and Mesiar R., 1999. Conditioning in possibility with strict order norms, Fuzzy Sets and Systems, 106, 221-229.

[14] De Cooman G., 1997. Possibility theory Part I: Measure- and integral-theoretics groundwork; Part II: Conditional possibility; Part III: Possibilistic independence, Int. J. of General Systems, 25(4), 291-371.

[15] De Cooman G., Aeyels D., 1999. Supremum-preserving upper probabilities. Information Sciences, 118, 173 -212.

[16] De Cooman G., 2001. Integration and conditioning in numerical possibility theory. Annals of Mathematics and AI, 32, 87-123

[17] De Cooman G., 2005. A behavioural model for vague probability assessments Fuzzy Sets and Systems, 154, 305-358.

[18] Delgado M. and Moral S., 1987. On the concept of possibility-probability consistency, Fuzzy Sets and Systems, 21, 311-318.

[19] Dempster A. P., 1967. Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat., 38, 325-339.

[20] Denneberg D., 1994. Nonadditive Measure and Integral, Kluwer Academic, Dordrecht, The Netherlands.

[21] Denoeux T., Masson M.-H., Hbert P.-A., 2005, Nonparametric rank-based statistics and significance tests for fuzzy data Fuzzy Sets and Systems, 153, 1-28.

[22] Dubois D., 1986. Belief structures, possibility theory and decomposable confidence measures on finite sets, Computers and Artificial Intelligence (Bratislava), 5, 403-416.

[23] Dubois D., Fargier H., Fortin J., 2005. The empirical variance of a set of fuzzy intervals. Proc. of the IEEE Int. Conf. on Fuzzy Systems, Reno, Nevada, IEEE Press, p. 885-890.

[24] Dubois D., Foulloy L., Mauris G., Prade H., 2004. Possibility/probability transformations, triangular fuzzy sets, and probabilistic inequalities. Reliable Computing, 10, 273-297.

[25] Dubois D. Huellermeier E., 2005. A notion of comparative probabilistic entropy based on the possibilistic specificity ordering. In L. Godo (Ed.) Proc. of the Europ. Conf. ECSQARU'05, Barcelona, LNAI 3571, Springer, Berlin, 848-859.

[26] Dubois D., Kerre E., Mesiar R., Prade H., 2000. Fuzzy interval analysis. In: D. Dubois H. Prade (Eds), Fundamentals of Fuzzy Sets, The Handbooks of Fuzzy Sets Series, Kluwer, Dordrecht, 483-581.

[27] Dubois D., Moral S. and Prade H., 1997. A semantics for possibility theory based on likelihoods, J. of Mathematical Analysis and Applications, 205, 359-380.

[28] Dubois D. and Prade H., 1980. Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York.

[29] Dubois and Prade H., 1982. On several representations of an uncertain body of evidence, In M.M. Gupta, and E. Sanchez (Eds) Fuzzy Information and Decision Processes, North-Holland, Amsterdam, 167-181.

[30] Dubois D. and Prade H., 1983. Unfair coins and necessity measures: towards a possibilistic interpretation of histograms. Fuzzy Sets and Systems, 10, 15-20.

[31] Dubois D. and Prade H., 1986. Fuzzy sets and statistical data, Europ. J. Operations Research, 25, 345-356.

[32] Dubois D. and Prade H., 1987. The mean value of a fuzzy number, Fuzzy Sets and Systems, 24, 279-300.

[33] Dubois D. and Prade H., 1988. Possibility Theory, Plenum Press, New York.

[34] Dubois D. and Prade H., 1990. Consonant approximations of belief functions. Int. J. Approximate Reasoning, 4, 419-449.

[35] Dubois D. and Prade H., 1991. Random sets and fuzzy interval analysis, Fuzzy Sets and Systems, 42, 87-101.

[36] Dubois D. and Prade H., 1992. When upper probabilities are possibility measures, Fuzzy Sets and Systems, 49, 65-74.

[37] Dubois D. and Prade H., 1997. Bayesian conditioning in possibility theory, Fuzzy Sets and Systems, 92, 223-240.

[38] Dubois D. and Prade H., 1998. Possibility theory: Qualitative and quantitative aspects, In Gabbay D.M. and Smets P., (eds.) Handbook of Defeasible Reasoning and Uncertainty Management Systems Vol. 1, Kluwer Academic Publ., Dordrecht, 169-226.

[39] Dubois D., Nguyen H. T., Prade H., 2000. Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps. In: D. Dubois H. Prade (Eds), Fundamentals of Fuzzy Sets, The Handbooks of Fuzzy Sets Series, Kluwer, Dordrecht, 343-438.

[40] Dubois D., Prade H., Rannou E., 1998. An improved method for finding typical values. Proc. of 7th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'98), Paris. Editions EDK, Paris, pp. 1830–1837.

[41] Dubois D., Prade H. and Sandri S., 1993. On possibility/probability transformations. In: Fuzzy Logic. State of the Art, In: R. Lowen, M. Roubens (eds.), Kluwer Acad. Publ., Dordrecht, 103-112.

[42] Dubois D., Prade H. and Smets P., 1996. Representing partial ignorance, IEEE Trans. on Systems, Man and Cybernetics, 26, 361-377.

[43] Dubois, D., Prade, H., Smets, P., 2001. Not impossible vs. guaranted possible in fusion and revision. Proc. of the 6th.European Conference (ESCQARU 2001), Toulouse, Springer-Verlag, LNAI 2143, 522-531

[44] Dubois D. Prade H., Smets, P., 2003. A definition of subjective possibility. Badania Operacyjne I Decyzje (Wroclaw University, Poland) #4, 7-22.

[45] D'Urso P., 2003. Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data.Computational Statistics & Data Analysis, 42, 47-72.

[46] Edwards W. F., 1972. Likelihood, Cambridge University Press, Cambridge, U.K.

[47] Ferson S., Ginzburg L.R., 1995. Hybrid Arithmetic. Proceedings of ISUMA-NAFIPS'95, IEEE Computer Society Press, 619-623.

[48] Ferson S., 1996. What Monte Carlo methods cannot do. Human and Ecology Risk Assessment, 2, 990-1007.

[49] Ferson S., Ginzburg L.R., 1996. Different methods are needed to propagate ignorance and variability. Reliability Engineering and Systems Safety, 54, 133-144.

[50] Ferson S., Ginzburg L., Kreinovich V., Longpre L., and Aviles M., 2002. Computing variance for interval data is NP-hard. ACM SIGACT News, 33, 108-118.

[51] Ferson, S., Berleant, D., Regan, H.M., 2004. Equivalence of methods for uncertainty propagation of real-valued random variables. International Journal of Approximate Reasoning, 36, 1-30

[52] Ferson S., Ginzburg L., Akcakaya R. 2006. Whereof one cannot speak: when input distributions are unknown. Risk Analysis, to appear.

[53] Fullér R., and Majlender P., 2003. On weighted possibilistic mean and variance of fuzzy numbersFuzzy Sets and Systems, 136,363-374.

[54] Fortemps P. and Roubens M., 1996. Ranking and defuzzification methods based on area compensation, Fuzzy Sets and Systems, 82, 319-330.

[55] Gebhardt J. and Kruse R., 1993. The context model: an integrating view of vagueness and uncertainty. Int. J. Approximate Reasoning, 9, 283-314.

[56] Gebhardt J. and Kruse R., 1994. A new approach to semantic aspects of possibilistic reasoning. In M. Clarke et al. (Eds.) Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Lecture Notes in Computer Sciences Vol. 747, Springer Verlag, 151-160.

[57] Gebhardt J. and Kruse R., 1994. On an information compression view of possibility theory. Proc 3rd IEEE Int. Conference on Fuzzy Systems. Orlando,Fl., 1285-1288.

[58] Geer J.F. and Klir G.J., 1992. A mathematical analysis of information-preserving transformations between probabilistic and possibilistic formulations of uncertainty, Int. J. of General Systems, 20, 143-176.

[59] Gil M. A., 1992. A note on the connection between fuzzy numbers and random intervals, Statistics and Probability Lett., 13, 311-319.

[60] Giles R., 1982. Foundations for a theory of possibility, Fuzzy Information and Decision Processes (Gupta M.M. and Sanchez E., eds.), North-Holland, 183-195.

[61] Guyonnet D., Bourgine B., Dubois D., Fargier H., Cume B., Chiles J.P., 2003 Hybrid approach for addressing uncertainty in risk assessments. Journal of Environmental Engineering, 126, 68-78.

[62] Kaufmann A. and Gupta M. M., 1985. Introduction to Fuzzy Arithmetic - Theory and Applications, Van Nostrand Reinhold, New York.

[63] Hacking I., 1975. All kinds of possibility, Philosophical Review, 84, 321-347.

[64] Halpern J., 2004. Reasoning about Uncertainty MIT Press, Cambridge, Mass.

[65] Hardy, G.H. Littlewood, J.E., Polya, G. Inequalities, Cambridge University Press, Cambridge UK, 1952.

[66] Heilpern S., 1992. The expected value of a fuzzy number, Fuzzy Sets and Systems, 47, 81-87.

[67] Heilpern S., 1997. Representation and application of fuzzy numbers, Fuzzy Sets and Systems, 91, 259-268.

[68] Higashi H. and Klir G., 1982. Measures of uncertainty and information based on possibility distributions, Int. J. General Systems, 8, 43-58.

[69] Hisdal E., 1978. Conditional possibilities independence and noninteraction, Fuzzy Sets and Systems, 1, 283-297.

[70] Hisdal E., 1988. Are grades of membership probabilities? Fuzzy Sets and Systems, 25, 325-348.

[71] Helton J.C. Oberkampf W.L., Eds., 2004. Alternative Representations of Uncertainty, Reliability Engineering and Systems Safety, vol. 85, Elsevier, 369 p.

[72] Hoffman F. O. and Hammonds J. S., 1994. Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability, Risk Analysis, vol. 14, pp. 707-712.

[73] Joslyn C., 1997. Measurement of possibilistic histograms from interval data, Int. J. of General Systems, 26(1-2), 9-33.

[74] Klir G.J., 1990. A principle of uncertainty and information invariance, Int. J. of General Systems, 17, 249-275.

[75] Klir G.J., 2006. Uncertainty and Information. Foundations of Generalized Information Theory. J. Wiley.

[76] Klir G.J. and Folger T., 1988. Fuzzy Sets, Uncertainty and Information, Prentice Hall, Englewood Cliffs, NJ.

[77] Klir G.J. and Parviz B., 1992. Probability-possibility transformations: A comparison, Int. J. of General Systems, 21, 291-310.

[78] Klir G.J., Ramer A., 1990. Uncertainty in the Dempster-Shafer theory : a critical reexamination. Int. J. of General Systems, 18(2), 155-166.

[79] Koerner R., 1997. On the variance of fuzzy random variable, Fuzzy Set and Systems, 92, 83-93.

[80] Krishnapuram R., Keller J., 1993. A possibilistic approach to clustering, IEEE Trans. Fuzzy Systems (1) 98-110.

[81] Lapointe S. Bobee B., 2000. Revision of possibility distributions: A Bayesian inference pattern, Fuzzy Sets and Systems, 116, 119-140

[82] Lindley D.V., 1982. Scoring rules and the inevitability of probability, Int. Statist. Rev., 50, 1-26.

[83] Mauris G., Lasserre V., Foulloy L., 2001. A fuzzy appproach for the expression of uncertainty in measurement. Int. J. Measurement, 29, 165-177.

[84] Levi I., 1973. Gambling with Truth, The MIT Press, Cambridge, Mass.

[85] Levi I., 1980. Potential surprize: its role in inference and decision-making. In: Applications of Inductive logic In L.J. Cohen, M. Hesse, (Eds), Oxford UP, UK.

[86] Lewis D. L., 1979. Counterfactuals and comparative possibility, In : Ifs, Harper W. L., Stalnaker R. and Pearce G., (eds.), D. Reidel, Dordrecht, 57-86.

[87] Masson, M. Denoeux T., 2006. Inferring a possibility distribution from empirical data, Fuzzy sets and Systems, 157, 319-340.

[88] H. Moulin. Axioms of Cooperative Decision Making. Cambridge University Press, Cambridge, MA, 1988.

[89] Neumaier A., 2004. Clouds, fuzzy sets and probability intervals. Reliable Computing, 10, 249-272.

[90] Nguyen, H. T., Bouchon-Meunier, B., 2003. Random sets and large deviations principle as a foundation for possibility measures, Soft Computing, 8:61-70.

[91] Nunez Garcia J., Kutalik Z. Cho K.-H and Wolkenhauer O., 2003. Level sets and minimum volume sets of probability density functions. Int. J. Approximate Reasoning, 34, 25-48.

[92] Ralescu D., 2002. Average level of a fuzzy set. In C. Bertoluzza, M. A. Gil, and D. A. Ralescu, (Eds), Statistical Modeling, Analysis and Management of Fuzzy data, Springer, Heidelberg, 119–126.

[93] Raufaste E., Da Silva Neves R., Mariné C., 2003. Testing the descriptive validity of possibility theory in human judgements of uncertainty. Artificial Intelligence, 148: 197-218.

[94] Saade J. J. and Schwarzlander H., 1992. Ordering fuzzy sets over the real line: An approach based on decision making under uncertainty, Fuzzy Sets and Systems, 50, 237-246.

[95] Shackle G. L.S., 1961. Decision, Order and Time in Human Affairs, (2nd edition), Cambridge University Press, UK.

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

[97] Shafer G., 1987. Belief functions and possibility measures, In Bezdek J.C., (Ed.) Analysis of Fuzzy Information Vol. I: Mathematics and Logic, CRC Press, Boca Raton, FL, 51-84.

[98] Shapley S., 1953. A value for n-person games. In Kuhn and Tucker, eds., Contributions to the Theory of Games, II, Princeton University Press, 307-317.

[99] Shilkret N., 1971. Maxitive measure and integration, Indag. Math., 33, 109-116.

[100] Smets P., 1990. Constructing the pignistic probability function in a context of uncertainty, In Henrion M. et al., (Eds.) Uncertainty in Artificial Intelligence, vol. 5, North-Holland, Amsterdam, 29-39.

[101] Smets P. and Kennes R., 1994. The transferable belief model, Artificial Intelligence, 66, 191-234.

[102] Spohn W., 1988. Ordinal conditional functions: A dynamic theory of epistemic states, In :Harper W. and Skyrms B., (Eds.) Causation in Decision, Belief Change and Statistics, 105-134.

[103] Tanaka, H., Uejima, S. and Asai, K., 1982. Linear regression analysis with fuzzy model. IEEE Trans. Systems Man Cybernet. 12, pp. 903-907.

[104] Tanaka, H., Lee H. 1999. Exponential possibility regression analysis by identification method of possibilistic coefficients Fuzzy Sets and Systems, 106, s 155-165.

[105] Thomas S.F., 1995. Fuzziness and Probability, ACG Press, Wichita, Kansas.

[106] Timm H., Borgelt C., Doering C., Kruse, R., 2004. An extension to possibilistic fuzzy cluster analysis. Fuzzy Sets and Systems, 147, 3-16.

[107] Van Leekwijk W. and Kerre E., 1999. Defuzzification: criteria and classification. Fuzzy Sets and Systems, 118, 159-178.

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

[109] Walley P., 1996. Measures of uncertainty in expert systems, Artificial Intelligence, 83, 1-58.

[110] Walley P. and de Cooman G., 1999. A behavioural model for linguistic uncertainty, Information Sciences, 134, 1-37.

[111] Walley P. and de Cooman G., 1999. Coherence of rules for defining conditional possibility, International Journal of Approximate Reasoning, 21, 63-107

[112] Wang P.Z., 1983. From the fuzzy statistics to the falling random subsets, In Wang P.P., (Ed.) Advances in Fuzzy Sets, Possibility Theory and Applications, Plenum Press, New York, 81-96.

[113] Yager R.R., 1980. A foundation for a theory of possibility, J. Cybernetics, 10, 177-204.

[114] Yager R. R., 1981. A procedure for ordering fuzzy subsets of the unit interval, Information Sciences, 24, 143-161.

[115] Yager R.R., 1992. On the specificity of a possibility distribution, Fuzzy Sets and Systems, 50, 279-292.

[116] Yager R. R. and Filev D., 1993. On the issue of defuzzification and selection based on a fuzzy set, Fuzzy Sets and Systems, 55, 255-271.

[117] Zadeh L.A., 1965. Fuzzy sets, Information and Control, 8, 338-353.

[118] Zadeh L. A., 1975. The concept of a linguistic variable and its application to approximate reasoning, Information Sciences, Part I: 8, 199-249; Part II: 8, 301-357; Part III: 9, 43-80.

[119] Zadeh L. A., 1978. Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems, 1, 3-28.