Uncertainty Theories: a Unified View

(Invited Paper, Cybernetic Systems Conference, Dublin (Ireland), 23-26/09/2007, IEEE, pp. 4-9, 2007)

Didier Dubois

***Abstract*—The variability of physical phenomena and partial ignorance about them motivated the development of probability theory in the two last centuries. However, the mathematical framework of probability theory, together with the Bayesian credo claiming the inevitability of unique probability measures for representing agents beliefs, have blurred the distinction between variability and ignorance. Modern theories of uncertainty, by putting together probabilistic and set-valued representations of information, provide a better account of the various facets of uncertainty.**

I. Introduction

The modeling of uncertainty is motivated by two concerns: taming the variability of observed phenomena and facing incomplete information in decision processes. These two concerns are related by the fact that in the face of variability, it is difficult to predict what the next state of the world will be. Yet the two concerns are distinct in the sense that variability is far from being the only cause of information incompleteness. Moreover, variability is an objective phenomenon, that supposedly corresponds to a property of the world. On the contrary, incompleteness of information refers to a human agent, and is thus irreducibly subjective. An agent cannot make prediction about the world behavior because of variability and because of lack of knowledge. But one may miss knowledge about quantities that are not subject to variability, like when guessing the birth date of a famous person, or the height of a mountain.

However, the development of probability theory, as witnessed by the Bayesian school especially, led to blur the distinction between variability and ignorance, suggesting that a unique probability distribution is enough to account for both randomness and incomplete information. More recently, new theories of uncertainty have emerged where partial ignorance is acknowledged and represented separately from randomness: the theories of imprecise probabilities, evidence, and possibility respectively. The aim of this paper is to suggest a unified view of these approaches.

II. Set-based representations of partial ignorance

The basic tool for representing information incompleteness is set theory: an ill-known quantity is represented as a disjunctive set, i.e. a subset of mutually exclusive values, one of which is the real one. This kind of uncertainty is naturally accounted for in logical representations. In propositional logic, a set of formulas viewed as constraints on the joint values of Boolean variables, generally does not single out a unique assignment (a model). When a given proposition is not entailed by a belief base, it does not imply that its negation is necessarily entailed: there are propositions, the truth of which cannot be inferred. This is when knowledge is incomplete.

Didier Dubois is with the Institut de Recherche en Informatique de Toulouse (IRIT), CNRS, Université de Toulouse, 31062 Toulouse Cedex 9, France (email: dubois@irit.fr).

In the area of numerical modelling, the processing of incomplete information is basically carried out by interval analysis [18] or constraint propagation methods. Incomplete information comes in the form of intervals assigned to unknown quantities. Assigning an interval $[a, b]$ to a quantity x means that x is known to take one and only one value in $[a, b]$, but it is nor known which one. Note that x is not necessarily a random quantity. It can be deterministic, yet unknown. For instance, the reader may not know the birth date of the president of Brazil, even if he or she can suggest a time interval. Note that such an interval like $[a, b]$ is never an attribute of the world, since the real value of x is precise. It is an attribute of an agent, an observer, a sensor, etc. In this sense, the set-valued representation is subjective: two individuals may come up with different intervals for the value of x, both of them being correct, even if not equally informative. The wider the set, the less informative. Total ignorance corresponds to the whole domain of x being possible.

The two modalities attached to this representation of partial ignorance are the possibility (or plausibility) and the certainty (or necessity), not probability. Asserting $x \in [a, b]$ comes down to declaring any value outside $[a, b]$ as impossible for x. Moreover, for an agent only knowing $x \in [a, b]$:

- Any event A understood as the assertion $x \in A$ is possible whenever $A \cap [a, b]$ is not empty.
- Any such event A is certain for this agent, whenever $[a, b] \subseteq A$.

It is clear that possibility and certainty correspond to logical consistency and logical entailment from the available knowledge. This is the Boolean version of possibility theory [12]. This type of uncertainty is captured by interval analysis. The basic problem in uncertainty analysis is: given a mathematical model of the form $y = f(x_1, \ldots, x_n)$, find the range of the output y, when all is known about inputs $x_i, i = 1, \ldots, n$ is that $x_i \in [a_i, b_i]$. This type of approach to partial ignorance, although perfectly sound, is crude and potentially little informative if intervals are too wide, each bound being respectively overpessimistic and overoptimistic.

III. Bayesian probability as a representation of partial ignorance

That probability theory may account for the variability of observed phenomena is clear, once statistical data is used to infer probabilities, interpreted as limit frequencies. However, this interpretation of probability is not valid for non-repeatable events. Several scholars in the XXth century, like Ramsey, De Finetti and Savage have introduced the notion of subjective probability so as to account for the fact that individuals entertain beliefs about non-repeatable situations. This trend has led to the Bayesian view of probability.

The operational definition of a degree of probability is then an amount of money an agent is ready to bet on the occurrence of an event. In such a betting experiment, the agent provides betting odds under an exchangeable bet assumption: it says that the agent is ready to buy a lottery ticket about the event at the same price as (s)he would sell it. On this basis, Bayesians claim that any state of incomplete knowledge of an agent can (and should) be modelled by a single probability distribution on the appropriate referential, and that degrees of belief coincide with probabilities that can be revealed by observing the betting behaviour of the agent. Failing to use a unique probability distribution, the agent is sure to lose money (this is called the Dutch book).

The idea that it is always possible to come up with a precise probability model, whatever the agent's state of knowledge, looks debatable. It is not clear that incomplete knowledge should be modelled by the same tool as variability (a unique probability distribution) [16]. One may argue, following e.g. Walley [24], 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. In the face of ignorance, it is not clear that individuals would buy and sell a lottery tickets at the same price: the selling price would be higher than the buying price, the difference being all the higher as the ignorance is significant. 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.

Bayesian betting rates cannot distinguish between the situation of total ignorance and the situation of complete knowledge about a perfectly random phenomenon. The same uniform probability accounts for an unknown die and a die known as fair. In the face of ignorance this uniform probability is enough for a decision-maker accepting Savage decision theory as the way to follow when choosing among acts. It becomes debatable when a new information requires a belief revision process or when conservative risk analysis must be carried out. In a crime case involving suspects Peter Paul and Mary, if it is only known that the killer is a man with probability one-half, a Bayesian may attach probability one fourth to Peter and to Paul. If it is known later that Peter has an alibi, it sounds safer to bet again on the updated set of suspects (at equal odds on Mary and Paul), than to consider Mary as having more chance to be the killer due to Bayes rule attaching probability two-third to her. In the case of risk analysis, based on propagating uncertainty through a mathematical model, precise subjective probability assessments of ill-known inputs will lead to alter the variance of the output in a debatable way: part of this calculated variance will be due to ignorance, not to effective variability, but they cannot be told apart. Yet, variance due to ignorance can be reduced by collecting more information, while variance due to variability just reflects the behaviour of the phenomenon under concern.

IV. Blending set-valued and probabilistic representations of uncertainty

Modern uncertainty theories put together probabilistic and set-valued representations, which allows for a clear separation between randomness and incompleteness. Moreover, it makes set-valued representations much more expressive. The most general approach consists in moving from the use of a single probability to a set of probabilities, all the larger as information is poor. It may express imprecision about an ill-known (objective) probability model and induces upper and lower probabilities of events. The subjective view can be accommodated by giving up the exchangeable bet assumption. The minimal selling and maximal buying prices of a lottery ticket winning one pound if an event occurs correspond to the upper and lower probabilities of the event, respectively [24]. Another way of blending probability and set representations is to randomize the latter. Random sets are the basis of the mathematical theory of evidence [22]. The idea is to define a probability distribution not on a state space, but on the powerset thereof. Each probability weight attached to a set corresponds to an amount of probability that should be shared among the elements of this set, but is not by lack of information. Upper and lower probabilities induced by this random set representation are special cases of imprecise probabilities. Finally, the last representation framework consists in assuming that possibility is a matter of degree, values in the set containing the unknown quantity being more or less plausible (or surprising [21]). This is possibility theory [13], based on possibility distributions. In possibility theory, information is thus summarized by fuzzy sets [25]. It is important to figure out that the reason why fuzzy set theory is relevant for uncertainty handling is because it is a set-based approach, not just because it is fuzzy in the sense of Zadeh (i.e. gradual rather than abrupt). The bridge to imprecise probability comes from the fact that fuzzy sets representing possibility distributions are equivalent to consonant (nested) random sets [13].

A. Imprecise probabilities

The theory of imprecise probabilities has been systematized and popularized by Walley's book [24]. In this theory, uncertainty is modeled by a family $\mathcal{P}$ of probability distributions. Lower and upper probability bounds are defined as follows :

$$P_*(A) = \inf_{P \in \mathcal{P}} P(A) \text{ and } P^*(A) = \sup_{P \in \mathcal{P}} P(A).$$

These two measures are dual to each other ($P_*(A) = 1 - P^*(A^c)$), and specifying one of them is enough to completely characterize the probability family. Let $\mathcal{P}_*^* = \{P | \forall A \subseteq X$ measurable, $P_*(A) \leq P(A) \leq P^*(A)\}$. In general, we have $\mathcal{P} \subset \mathcal{P}_*^*$, since $\mathcal{P}_*^*$ can be seen as a projection of $\mathcal{P}$ on events. A family $\mathcal{P}$ can also be defined by a set of restrictions of the type

$$\underline{P}(A) \leq \sum_{x_i \in A} p(x_i) \leq \overline{P}(A)$$

When it defines a non empty probability set $\mathcal{P}$ and each bound $\underline{P}(A)$ and $\overline{P}(A)$ is attained by one probability measure in $\mathcal{P}$, the representation is said to be coherent.

B. Random disjunctive sets

Formally, a random set is a mapping from a probability space $(\Omega, \mathcal{A}, P)$ to the power set $\wp(X)$ of another space X, also called a multi-valued mapping Γ. The set $\Gamma(\omega)$ represent incomplete knowledge about a random variable when the realization is ω in the probability space (as opposed to random sets which are conjoints of elements like an ill-known region in a digital image). Then this multimapping induces a probability family on X representing all probability functions on X that could be found from all measurable mappings $\Omega \to X$ compatible with Γ [8]. Upper and lower probabilities probabilities on events in X are then generated. Such lower and upper probabilities are respectively called belief and plausibility functions by Shafer [22] and denoted Bel and Pl respectively. He uses an alternative (and useful) representation of the random set consisting of a distribution of positive masses m over the power set $\wp(X)$ s.t. $\sum_{E \subseteq X} m(E) = 1$ and $m(\emptyset) = 0$. Namely if $E = \Gamma(\omega)$, then let $m(E) = p(\omega)$. A set E that receives strictly positive mass is called a focal set. The mass $m(E)$ is interpreted as the probability of knowing only E as containing the actual solution to the problem under concern. We have :

$$\begin{aligned} Bel(A) &= \sum_{E, E \subseteq A} m(E) \\ Pl(A) &= 1 - Bel(A^c) \\ &= \sum_{E, E \cap A \neq \emptyset} m(E). \end{aligned}$$

There is a one-to-one correspondence between the mass distribution and the belief function since

$$m(E) = \sum_{B \subset E} (-1)^{|E-B|} Bel(B).$$

This last equation is known as the Möbius inverse and can be applied to any kind of lower probability function. The positivity of the mass function obtained by the Möbius inverse is characteristic of the random set setting. In the finite case, it can be shown that the lower probability function is an ∞-monotone capacity. The set $\mathcal{P}_{Bel} = \{P | \forall A \subseteq X$ measurable, $Bel(A) \leq P(A) \leq Pl(A)\}$ is coherent and forms the probability family induced by the belief function. Note that Shafer [22] does not refer to an underlying probability space, nor does he uses the fact that a belief function is a lower probability: in his view, extensively taken over by Smets [23], $Bel(A)$ is supposed to quantify an agent's belief per se with no reference to a probability. However, the basic mathematical tool common to Dempster's upper and lower probabilities and to the Shafer-Smets view is the notion of (generally finite) random disjunctive set.

C. Fuzzy sets as possibility distributions

A possibility distribution π [12] is a mapping from X to the unit interval such that $\pi(x) = 1$ for some $x \in X$. Formally, a possibility distribution is equivalent to the membership function of fuzzy set $\mu(x) = \pi(x)\ \forall x$. Possibility distributions were introduced by Zadeh [25] as flexible constraints induced by fuzzy natural language statements. Twenty years earlier, Shackle [21] had introduced an equivalent notion called distribution of potential surprise (corresponding to $1 - \pi(x)$) with a view to represent non-probabilistic uncertainty. Several set-functions can be defined from a possibility distribution π, especially [12]:

- Possibility measures: $\Pi(A) = \sup_{x \in A} \pi(x)$
- Necessity measures: $N(A) = 1 - \Pi(A^c)$

The possibility degree of an event expresses the extent to which this event is plausible, i.e., consistent with a possible state of the world. Necessity degrees express the certainty of events, by duality. Under these measures, the possibility quantified by distribution π is potential (in the spirit of Shackle), i.e. $\pi(x)$ = 1 does not guarantee the corresponding value x.

A (potential) possibility degree can be viewed as an upper bound of a probability degree [13]. Let $\mathcal{P}_\pi = \{P, \forall A \subseteq X$ measurable, $N(A) \leq P(A) \leq \Pi(A)\}$ be the set of probability measures encoded by a possibility distribution π. This representation is coherent since upper and lower probabilities induced by $\mathcal{P}_\pi$ are precisely Π and N.

In the finite case, a possibility distribution is also equivalent to a random set whose focal elements are nested. Namely, a belief function (resp. a plausibility function) is a necessity measure (resp a possibility measure) if and only if they derive from a mass function with nested focal sets (already in [22]). Their characteristic property is then $N(A \cap B) = \min(N(A), N(B))$ (resp. $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$. In this situation, the same amount of information is contained in the mass function m and the possibility distribution $\pi(x) = Pl(\{x\})$. In the general non-nested case, m cannot be reconstructed from $Pl(\{x\})$.

V. Practical representations of incomplete knowledge

Imprecise probabilities are complex to represent, much more than probability measures, whether on finite sets or the real line. On a finite set X with N elements, a lower probability needs 2^N values to be specified (and a consistency check to make it sure that the corresponding family is not empty). It defines a convex polyhedron, which may have up to $N!$ vertices. If X is defined via a set of Boolean variables, there is a counterpart of Bayesian networks, called credal networks, which allow the use of probability bounds on much smaller subspaces. In general, to completely specify a probability family induced by a random set, one still needs to give $2^{|X|}$ different values, thus not reducing the complexity of the representation with respect to a capacity. However, simple belief functions having only a few positive focal sets do not exhibit such a complexity. From a computational perspective, the main advantage of random sets is that they can be seen as a probability distribution over subsets of X. Therefore, they can easily be simulated by some process such as Monte-Carlo sampling, which is not the case for other Choquet capacities. On the real line, a random set is often restricted to a finite collection of closed intervals with associated weights, and one can then easily extend results from interval analysis [18] to random intervals [20]. For continuous random intervals, the mass function is replaced by a mass density bearing on intervals. In the following we point out existing simple representations of probability families [9].

A. Fuzzy sets and intervals

At most $|X|$ values are needed to completely specify a possibility distribution, making them easier to represent than general random sets. On the real line, continuous (or upper semi-continuous) unimodal possibility distributions on the real line encompass closed intervals and are called fuzzy intervals [12]. They have a very natural interpretation as sets of nested confidence intervals [10] or probabilistic inequalities, like Chebyshev's [11]. Ill-known probability models where only some parameters are known, like the support and the mode, are liable to a possibilistic representation [1]. Possibility distributions are thus the simplest models of probability families, and they play a central role when modeling vague assessments of probabilities (see [6]).

B. P-boxes and generalized p-boxes

A p-box [17] is defined by a pair of cumulative distributions on the real line such that $\underline{F} \leq \overline{F}$, bounding the cumulative distribution of an imprecisely known probability function with density p. It can be viewed as a generalized interval as well. The corresponding set of probabilities $\mathcal{P}_{p-box}$ is representable by a belief function whose focal elements are of the form $\{x, \overline{F}(x) \geq \alpha\} \setminus \{x, \underline{F}(x) \geq \alpha\}$. A p-box is a covering approximation of a parameterized probability model whose parameters (like mean and variance) are only known to belong to an interval.

A p-box can be induced from a possibility distribution π, letting $\underline{F}(x) = N((-\infty, x])$ and $\overline{F}(x) = \Pi((-\infty, x])$, but the probability family $\mathcal{P}_{p-box}$ induced by this p-box is strictly larger than $\mathcal{P}_\pi$[1]. So, while a fuzzy interval induces a p-box, such generated p-boxes are less informative than the possibility distributions they are computed from.

Interestingly, the notion of cumulative distribution is based on the existence of the natural ordering of numbers. On a finite set, no obvious notion of cumulative distribution exists. In order to make sense of this notion over X, one must equip it with a complete preordering and define two cumulative distributions according to this ordering. it comes down to a family of nested confidence sets $\emptyset \subseteq A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subset X$. The family $\mathcal{P}_{p-box}$ can then be represented by the following restrictions on probability measures

$$\alpha_i \leq P(A_i) \leq \beta_i \qquad i = 1, \ldots, n \tag{1}$$

with $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n \leq 1$ and $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_n \leq 1$. If we take $X = \Re$ and $A_i = (-\infty, x_i]$, it is easy to see that we retrieve the usual definition of P-boxes.

A family $\mathcal{P}_{p-box}$ described by a generalized P-box can be encoded by a pair of possibility distributions π_1, π_2 s.t. $\mathcal{P}_{p-box} = \mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2}$ where π_1 comes from constraints $\alpha_i \leq P(A_i)$ and π_2 from constraints $P(A_i) \leq \beta_i$. Again, it is representable by a belief function [9].

C. Probability Intervals

Probability intervals over the finite space X are defined as lower and upper probability bounds restricted to singletons x_i [5]. They can be seen as a set of intervals $L = \{[l_i, u_i], i = 1, \ldots, n\}$ defining the family

$$\mathcal{P}_L = \{P | l_i \leq p(x_i) \leq u_i \forall x_i \in X\}$$

it is easy to see that $\mathcal{P}_L$ is totally determined by $2|X|$ values. $\mathcal{P}_L$ is non-empty provided that $\sum_{i=1}^n l_i \leq 1 \leq \sum_{i=1}^n u_i$.

A set of probability intervals L will be called reachable if, for each x_i, each bound u_i and l_i can be reached by at least one distribution of the family $\mathcal{P}_L$. Reachability is equivalent to the condition

$$\sum_{j \neq i} l_j + u_i \leq 1 \text{ and } \sum_{j \neq i} u_j + l_i \geq 1$$

Lower and upper probabilities $P_*(A), P^*(A)$ are calculated by the following expressions

$$\begin{aligned} P_*(A) = &\quad \max(\textstyle\sum_{x_i \in A} l_i, 1 - \sum_{x_i \notin A} u_i), \\ P^*(A) = &\quad \min(\textstyle\sum_{x_i \in A} u_i, 1 - \sum_{x_i \notin A} l_i). \end{aligned}$$

De Campos et al. [5] have shown that these bounds are Choquet capacities of order 2. It is easy to extract p-boxes from probability intervals and conversely, but some information is lost in the process.

D. Clouds

A *cloud* [19] can be seen as an Interval-Valued Fuzzy Set F such that $(0, 1) \subseteq \cup_{x \in X} F(x) \subseteq [0, 1]$, where $F(x)$ is an interval $[\delta(x), \pi(x)]$. It implies that $\pi(x) = 1$ for some x and $\delta(y) = 0$ for some y. A probability measure P on X is said to belong to a cloud F if and only if $\forall \alpha \in [0, 1]$:

$$P(\delta(x) \geq \alpha) \leq 1 - \alpha \leq P(\pi(x) > \alpha) \tag{2}$$

under all suitable measurability assumptions. From this definition, a cloud (δ, π) is equivalent to the cloud $(1 - \pi, 1 - \delta)$. If X is a finite space of cardinality n, let $A_i = \{x_i, \pi(x_i) >$

$\alpha_{i+1}\}$ and $B_i = \{x_i, \delta(x_i) \geq \alpha_{i+1}\}$ A cloud can thus be defined by the following restrictions [9]:

$$P(B_i) \leq 1 - \alpha_i \leq P(A_i) \text{ and } B_i \subseteq A_i \quad i = 1, \ldots, n \quad (3)$$

where $1 = \alpha_0 > \alpha_1 > \alpha_2 > \ldots > \alpha_n > \alpha_{n+1} = 0$ and $\emptyset = A_0 \subset A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq A_{n+1} = X; \emptyset = B_0 \subseteq B_1 \subseteq B_2 \subseteq \ldots \subseteq B_n \subseteq B_{n+1} = X$.

Let $\mathcal{P}_{\delta,\pi}$ be the probability family described by the cloud (δ, π) on a referential X. Clouds are closely related to possibility distributions and p-boxes as follows [9]:

- $\mathcal{P}_{\delta,\pi} = \mathcal{P}_\pi \cap \mathcal{P}_{1-\delta}$ using the probability families induced by the two possibility distributions π and $1 - \delta$.
- A cloud is a generalized p-box iff the sets $\{A_i, B_i, i = 1, \ldots, n\}$ form a nested sequence (i.e. there is a complete order with respect to inclusion); in other words, it means that π and δ are comonotonic. So a comonotonic cloud generates upper and lower probabilities that are plausibility and belief functions.
- When the cloud is not comonotonic, $\mathcal{P}_{\delta,\pi}$ generates lower probabilities that are not even 2-monotone. It is anyway possible to approximate upper and lower probabilities of events from the outside by possibility and necessity measures based on π and $1 - \delta$:

$$\max(N_\pi(A), N_{1-\delta}(A)) \leq P(A) \leq \min(\Pi_\pi(A), \Pi_{1-\delta}(A)).$$

The belief and plausibility functions of the random set s.t. $m(A_i \setminus B_{i-1}) = \alpha_{i-1} - \alpha_i$ are inner approximations of $\mathcal{P}_{\delta,\pi}$, which become exact when the cloud is monotonic.

When $\pi = \delta$ the cloud is said to be thin. In the finite case, $\mathcal{P}_{\pi,\pi} = \emptyset$. To make it not empty, we need a one-step index shift, such that (assuming the π_i's are decreasingly ordered) $\delta_i = 1 - \pi_{i+1}$ (with $\pi_{n+1} = 0$). Then, $\mathcal{P}_{\delta,\pi}$ contains a single distribution p such that $p_i = \pi_i - \pi_{i+1}$. In the continuous case $\mathcal{P}_{\pi,\pi}$ contains an infinity of probability measures and corresponds to a random set whose realizations are doubletons (the end-points of the cuts of π).

The above representations of imprecise probabilities are easier to handle than general probability families. They often require less evaluations to be fully specified and they allow many mathematical simplifications that increase computational efficiency (except, perhaps, for non-comonotonic clouds).

VI. Basic problems in uncertainty theories

Probability theory is used either to capture stochastic features of a population of situations, or to measure degrees of belief of agents. Even if the two purposes converge for agents that derive their degrees of belief from the knowledge of frequencies, the two kinds of probability are not of the same nature. While frequentist probabilities refer to a population of situations and represent generic knowledge, subjective probabilities are elicited on the basis of single contingent events. This distinction is important to better understand various uncertainty management problems and how to to state and solve them when imprecise probabilities enter the picture. Especially, insofar as one may admit that reality is precise but only our perception of reality is imprecise, there cannot be such thing as a fully objective imprecise probability model. Even when capturing variability, i.e. when a set $\mathcal{P}$ of probability measures is assumed to contain the "true" probability measure P that governs a process, the imprecision itself is always subjective in the sense that it will vary from one observer to another. As for subjective upper and lower probabilities, they can be viewed as degrees of belief and plausibility of contingent events, without assuming the existence of a true probabilistic representation of belief (both Shafer [22] and Smets [23] reject this assumption) and without reference to statistics. Shafer-Smets evidence theory is about uncertain testimonies, not about induction from data tainted with variability.

In the following, we briefly review basic problems of imprecise probability management:

- *Learning imprecise models vs. eliciting degrees of belief* Since the assumption of a unique probability distribution is often made in statistics, the basic learning problem is how to is to extract a probability distribution from data. When data is scarce, it would seem more natural to get imprecise probabilities. For instance, confidence intervals on distribution parameters naturally define probability families. Moreover the Imprecise Dirichlet model [3] offers a tool to relate the number of observations to the width of probability intervals without resorting to so-called uninformative priors. Another important problem is to adapt classical methods for inferential statistics to the case of imprecise (interval or fuzzy) data. In the case of subjective probabilities, elicitation techniques are tailored to produce unique probability functions. Yet expert knowledge is imprecise, and more naturally comes under the form of nested intervals with various confidence levels. Some effort is needed to adapt existing elicitation techniques towards producing possibility distributions and generalized p-boxes, that naturally capture nested intervals.
- *Uncertainty propagation* Given several inputs, some modeled by intervals, some modeled by probability distributions, some by possibility distributions, how to compute the output of a mathematical models? It is possible to use Monte-Carlo methods conjoined with interval analysis, when the imprecise probabilities derive from random sets [20], [2]. However the important issue of modeling independence and accounting for partial knowledge about dependence in imprecise probability theories remains an open topic to a large extent [4].
- *Inference from contingent information* A classical use of probabilistic models is the following: given a probabilistic model, representing generic knowledge, and deterministic observations on a contingent situation, predict whether some property is true in this situation (diagnosing a fault, assigning a class to an instance, etc...). This is done by Bayesian conditioning. In the case of imprecise probability models, this is done by an extension of Bayesian conditioning consisting in finding

upper and lower conditional probabilities $P(A \mid B)$, where B captures the contingent observations, and A is the event of interest, when P varies in $\{P \in \mathcal{P}, P(B) > 0\}$ [24]. This is a matter of querying the knowledge $\mathcal{P}$ by focusing it to the reference class B [15].

- *Revision* In the above problem, the set of probabilities $\mathcal{P}$ does not evolve because it is not of the same nature as the contingent information. In the case of revision. $\mathcal{P}$ represents prior incomplete uncertain information about a case, and B is of the same nature, yet certain. The input information tells us that $P(B) = 1$, so that some subjective probabilities in $\mathcal{P}$ can be ruled out. It leads to changing $\mathcal{P}$ into $\{P \in \mathcal{P}, P(B) = 1\}$, or, if empty, $\{P \in \mathcal{P}, P(B)$ maximal $\}$. Then, in many cases, $P^*(A \mid B) = \frac{P^*(A \cap B)}{P^*(A)}$ [15]. In evidence theory, this is Dempster rule of conditioning [22]. It revises $\mathcal{P}$ and differs from the imprecise Bayesian conditioning rule, even if both rules coincide if $\mathcal{P}$ contains a single probability measure.
- *Fusion* The fusion problem consists in merging several bodies of uncertain information of the same nature issued from several sources. This is a problem that was of interest at the origins of probability theory (the merging of uncertain testimonies in legal procedures). It was revived in the late XXth century with the emergence of the computer, in robotics (merging sensor information), in reliability (merging expert opinions), in signal processing (merging digital images), etc. There are now many fusion rules especially in possibility theory [14], evidence theory [22], and in probability theory as well [7]. Fusion can be viewed as a preliminary step before inference (laying bare a 'sure input" from contradictory reports), and as a generalization of revision (Dempster rule of combination [22] can be viewed as a generalization of Dempster rule of conditioning).
- *Statistics with imprecise data* In the face of imprecise or fuzzy data, or for the purpose of summarizing the output of an uncertainty propagation or a fusion procedure in the framework of imprecise probabilities, it is useful to summarize information for the user. It is possible to compute mean values and variances, empirical or theoretical; however they will generally be imprecise, and can be difficult to compute. P-boxes can also be extracted [2]. But such p-boxes are generally not enough to address questions such as the probability for an output to lie between two bounds. A considerable effort is required to make imprecise statistics palatable for users.

VII. Conclusion

Imprecise probabilities are a natural concept for conjointly handling variability of phenomena and incomplete knowledge about them. Imprecise modeling is unusual. In classical approaches, a probabilistic model is an approximate but precise representation of variability. It is assumed that the precise results it produces are not too far from observations of a random reality. In contrast, an imprecise model is of higher order: it represents both knowledge about reality and knowledge about knowledge. There is a need to reconsider the foundations of statistics in this perspective. We reviewed simplified representations of imprecise probabilities, that should eventually help solving related computational problems at the practical level.

Acknowledgment

This paper is partially supported by the ANR project CRISCO2. It owes much to my colleagues and students, especially C. Baudrit, S. Destercke and H. Prade.

References

[1] C. Baudrit and D. Dubois Practical representations of incomplete probabilistic knowledge Computational Statistics & Data Analysis. 51, 86-108, 2006.

[2] C. Baudrit D. Guyonnet, and D. Dubois Joint Propagation and Exploitation of Probabilistic and Possibilistic Information in Risk Assessment, IEEE Trans. Fuzzy Systems, 14, 593-608, 2006

[3] J.M. Bernard, An introduction to the imprecise Dirichlet model for multinomial data. J. Approximate Reasoning, 39, 123-150, 2005

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

[5] L.M. Campos and J.F. Huete and S. Moral, Probability intervals : a tool for uncertain reasoning. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, 2, 167-196, 1994.

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

[7] R. Cooke, Experts in Uncertainty, Oxford UP, 1991.

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

[9] S. Destercke, D. Dubois, E. Chojnacki,On the relations between practical representations of imprecise probabilities, ISPITA Conference, Prague, 2007.

[10] D. Dubois, Possibility theory and statistical reasoning. Computational Statistics & Data Analysis, 51, 47-69, 2006.

[11] D. Dubois, L. Foulloy, G. Mauris, and H. Prade, Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. Reliable Computing, 10, 273–297, 2004.

[12] D. Dubois and H. Prade, Possibility Theory, New York: Plenum, 1988.

[13] D. Dubois, H.T. Nguyen and H. Prade, Fuzzy sets and probability: misunderstandings, bridges and gaps. In D. Dubois and H. Prade, eds, Fundamentals of Fuzzy Sets. Boston, Mass: Kluwer, 343-438, 2000.

[14] D. Dubois, H. Prade, Possibility theory in information fusion. In: Data Fusion and Perception. G. Della Riccia et al., Eds, CISM Series V. 431, Springer-Verlag, 2002, 53-76.

[15] D. Dubois, H. Prade, P. Smets, Representing partial ignorance. IEEE Trans. on Systems, Man and Cybernetics, 26(3), 1996, 361-377.

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

[17] S. Ferson et al., Construction of probability boxes and Dempster-Shafer structures, Sandia National Laboratories, Technical report SANDD2002-4015, 2003.

[18] R.E. Moore Methods and applications of Interval Analysis, SIAM Studies in Applied Mathematics, 1979

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

[20] W.L. Oberkampf et al., Challenge problems : uncertainty in system response given uncertain parameters. Reliability Engineering and System Safety, 85, 11-19, 2004.

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

[22] G. Shafer, A mathematical Theory of Evidence, Princeton University Press, 1976.

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

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

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