Formal representations of uncertainty

Didier Dubois, Henri Prade

September 21, 2008

1 Introduction

The recent development of uncertainty theories that account for the notion of belief is linked to the emergence, in the XXth century, of Decision Theory and Artificial Intelligence. Nevertheless, this topic was dealt with very differently by each area. Decision Theory insisted on the necessity to found representations on the empirical observation of individuals choosing between courses of action, regardless of any other type of information. Any axiom in the theory should be liable of empirical validation. Probabilistic representations of uncertainty can then be justified with a subjectivist point of view, without necessary reference to frequency. Degrees of probability then evaluate to what extent an agent believes in the occurrence of an event or in the truth of a proposition. In contrast, Artificial Intelligence adopted a more introspective approach aiming at formalizing intuitions, reasoning processes, through the statement of reasonable axioms, often without reference to probability. Actually, until the nineties Artificial Intelligence essentially focused on purely qualitative and ordinal (in fact, logical) representations.

Historically, the interest for formalizing uncertainty appears in the middle of the XVIIth century, involving scholars like Pascal, Fermat, Huyghens, the chevalier de Méré, Jacob Bernoulli, etc. Two distinct notions were laid bare and studied: the objective notion of chance, related to the analysis of games, and the subjective notion of probability in connection to the issue of the reliability of witnesses in legal matters. In pioneering works like those of J. Bernoulli, chances were quickly related to the evaluation of frequency and thus naturally additive, but probabilities were not considered so in the first stand. However, in the XVIIIth century, with the fast developments of hard sciences, the interest in the subjective side of probability waned and the additive side of probability became prominent, so much so as some late works on non-additive probabilities (for instance those of Lambert) became unpalatable to other contemporaneous scholars (see Shafer [105]). From then on, under the influence of Laplace, and for long, probabilities would be additive, whether frequentist or not. It was thus very natural that in the XXth century, pioneering proposals formalizing probability independently from frequency (Ramsey [98], De Finetti [23]) tried to justify the additivity axiom, in the framework of a theory of gambles for economic decision problems, especially.

The raise of computer and information sciences in the last part of the XXth century renewed the interest for human knowledge representation and reasoning often tainted with imprecision, uncertainty, contradictions, independently of progress made in probability and decision theories, and focusing on the contrary on qualitative logical formalisms, especially in Artificial Intelligence, as well as the representation of the gradual nature of linguistic information (especially, fuzzy set theory). This trend has equally triggered the revival of non-additive probabilities for modelling uncertainty, a revival already pioneered by the works of Good [70], Smith [117], Shackle [103], Dempster [25], Kyburg [83], and Shafer [104]. Besides, the logic school rediscovered ancient modal concepts of possibility and necessity, quite relevant for epistemic issues, introduced by Aristotle and exploited by medieval religious philosophy. At the heart of the logical approach, the idea of incomplete knowledge is basic, and comes close to issues in imprecise probability (as opposed to the use of a unique probability distribution advocated by the Bayesian school). In the imprecise probability view, possibility and necessity respectively formalise subjective plausibility and certainty by means of upper and lower probability bounds. Such non-classical probabilistic independently appeared later on within Decision Theory itself, due to the questioning of the empirical validity of Savage's postulates underlying expected utility theory (Schmeidler [102]), after observing systematic violations of some of these postulates.

The gap created in the early XXith century between logicians (mainly interested by the foundation of mathematics) and statisticians now tends to reduce. To-date logic in its classical and non-classical versions (modal, non-monotonic, probabilistic, possibilistic) is again considered as a formal tool for the representation of human knowledge and the mechanization of reasoning processes, and is no longer confined to metamathematics. Along this line, it sounds more natural to propose that when statistical data is missing, the probabilistic knowledge possessed by an individual be represented by a set of logical propositions each having its probability, rather than by a probability distribution over an exhaustive set of mutually exclusive elements. However the former representation generally characterize a family of probability functions and not a unique distribution. This logical view of probability is present in the XIXth century, in the works of Boole [6], whose magnum opus *the Laws of Thought* lays the formal foundations of probabilistic logic at least as much as those of classical logic. Besides, Artificial Intelligence and Cognitive Psychology do share the concern of studying the laws of thought (even if with totally different goals).

The aim of this chapter is to propose a unified overview of various approaches to representations of uncertainty that came to light in the last fifty years or so in the areas of Artificial Intelligence and Decision Theory. The focus is on ideas and intuitions rather than on mathematical details. It is pointed out that apart from the central issue of belief representation, other aspects of the imperfection of information are currently studied for their own sake, such as the non-Boolean nature of linguistic predicates, and the concept of granularity. This chapter is organized as follows: the next section considers the notion of information in its semantic side and proposes a typology of defects of information items possessed by a cognitive entity (a human agent or a computer). Section 3 recalls some basics of probability theory, which in any case stands as a landmark. Injecting incomplete information into probability theory leads to a hierarchy of representations involving convex sets of probabilities, including Shafer's theory of evidence [104] and the numerical variant of possibility theory [43, 38]. These approaches are reviewed in section 4. This section also discusses bridges between possibility and probability. It is shown that some results and methods in non-Bayesian statistics can be reinterpreted and systematized in possibility theory, such as the maximum likelihood principle and confidence intervals. Moreover, the insufficient reason principle of Laplace can be extended to derive a probability measure from a possibility measure or conversely so as to justify possibility distributions as cautious substitutes of subjective probability distributions. Section 5 presents ordinal and logical representations of uncertainty. Qualitative possibility theory

[51] is tailored to handle incomplete information and is shown to stand as the simplest among ordinal approaches to uncertainty. Section 6 discusses the important notion of conditioning in uncertainty theories, using the key-concept of conditional event as a guideline. The bottom line is that, in probability theory, Bayesian conditioning is a unique tool instrumental for several distinct problems, but each problem requires a specific conditioning tool in the non-additive frameworks. Finally Section 7 deals with uncertain fusion information, and shows that the framework of uncertainty theories leaving room to incompleteness leads to a reconciliation of probabilistic fusion modes (based on averaging) and logical ones (based on conjunction and disjunction).

2 Information : a typology of defects

The term *information* refers to any collection of symbols or signs produced either through the observation of natural or artificial phenomena or by cognitive human activity with a view to help an agent understand the world or the current situation, making decisions, or communicating with other human or artificial agents. In this paper we focus on the mathematical representation of information items. We draw several important distinctions so as to charter this area.

A first distinction separates so-called *objective* information stemming from sensor measurements and the direct perception of events from *subjective* information typically uttered by individuals (e.g. testimonies) or conceived without resorting to direct observations.

Another distinction is between *quantitative* information modelled in terms of numbers, typically objective information (sensor measurements, counting processes), and *qualitative* or symbolic information (typically subjective information, e.g. expressed in natural language). Nevertheless, this partition is not so strict as it looks: subjective information can be numerical, and objective information can be qualitative (a color identified by means of a symbolic sensor, for instance). Quantitative information can assume various formats : numbers, intervals, functions. Structured symbolic information is often encoded in logical or graphical representations. There are hybrid representations such as weighted logics or probabilistic networks.

Yet another very important distinction must be drawn between singular and generic information. *Singular* information refers to a particular situation, a response to a question on the current state of affairs : for instance, an observation (a patient has fever at a given time point), or a testimony (the crazy driver's car was blue). *Generic* information refers to a collection or a population of situations (it could be a physical law, a statistical model built from a representative sample of observations, or yet a piece of commonsense knowledge like "birds fly"). This distinction is important when considering problems of inference or revision of uncertain information. Moreover, topic like induction or learning processes, deal with the construction of generic knowledge from several items of singular information. Conversely statistical prediction can be viewed as the use of some piece of generic knowledge on the frequency of an event to derive a degree of belief in the singular occurrence of this event in a specific situation (Hacking [72]).

An agent is supposed to have some information about the current world. The *epistemic state* of the agent is supposed to be made of three components: her generic knowledge, her singular observations, and her beliefs [34]. Beliefs are understood as pertaining to the current situation.

They are singular and derived from the two former kinds of information. They are instrumental to make decisions. Decision-making involves another kind of information possessed by an agent, this chapter does not deal with: her preferences.

In order to represent the epistemic state of an agent, a representation of the states of the world is needed, in agreement with the point of view of this agent, i.e., highlighting the relevant aspects by means of suitable attributes. Let v be the vector of attributes relevant for the agent and S the domain of v. S is called a *frame*; it is the set of (descriptions of) all states of the world. A subset A of S, viewed as a disjunction of possible worlds, is called an *event*, to be seen as a proposition that asserts $v \in A$. It is not supposed that the set S be explicitly known as a primitive object. It can as well be reconstructed, at least partially, from pieces of information supplied by the agent in the form of asserted propositions.

Four kinds of qualification of the imperfection of pieces of information expressible on the frame S can be considered : incomplete (or yet imprecise), uncertain, gradual, and granular information.

2.1 Incompleteness and Imprecision

A piece of information is said to be *incomplete* in a given context if it is not sufficient to allow the agent to answer a relevant question in this context. We interpret imprecision as a form of incompleteness, in the sense that an imprecise response provides only incomplete information. A kind of question to which the agent tries to answer is of the form *what is the current value of some quantity* v*?* or more generally : *does* v *satisfy some property of interest?* The notion of imprecision is not an absolute one. For instance, if the quantity of concern is the age of a person, the term *minor* is precise if the proper frame is $S = \{minor, major\}$ and the question of interest is : can the person vote? On the contrary, if $S = \{0, 1, \ldots, 150\}$ (in years), the term *minor* is imprecise, it provides incomplete information if the question of interest is to know the birth date of the person.

The typical form of a piece of incomplete information is $v \in A$ where A is a subset of S containing more than one element. An important remark is that elements in A, seen as possible values of v are mutually exclusive (since the quantity takes on a single value). Hence, a piece of imprecise information takes the form of a disjunction of mutually exclusive values. For instance, to say that *Pierre is between 20 and 25 years old*, i.e., $v =$*age(Pierre)* $\in \{20, 21, 22, 23, 24, 25\}$, is to suppose $v = 20$ or $v = 21$ or $v = 22$ or $v = 23$ or $v = 24$ or $v = 25$. In classical logic, incompleteness explicitly appears as a disjunction. Asserting the truth of $p \vee q$, means that one of the following propositions $p \wedge q$, $p \wedge \neg q$, $\neg p \wedge q$, is true. More generally, one of the models of $p \vee q$ is true.

A set used for representing a piece of incomplete information is called a *disjunctive set*. It contrasts with the conjunctive view of a set considered as a collection of elements. A conjunctive set represents a precise piece of information. For instance, consider the quantity $v = sisters(Pierre)$ whose range is the set of subsets of possible names for Pierre's sisters. The piece of information $v = \{Marie, Sylvie\}$ is precise and means that Pierre's sisters are Marie *and* Sylvie. Indeed, the frame is then $S = 2^{NAMES}$, where $NAMES$ is the set of all female first names. In this setting, a piece of incomplete information would be encoded as a disjunction of conjunctive subsets of $NAMES$.

A piece of incomplete information defines a so-called *possibility distribution* on S. If the available information is of the form $v \in A$, it means that any value of v not in A is considered impossible, but any value of v in the set A is possible. The possibility distribution encoding the piece of information $v \in A$, denoted by π_v is the characteristic function of A. It is a mapping from S to $\{0, 1\}$ such that $\pi_v(s) = 1$ if $s \in A$, and 0 otherwise. Conventions for $\pi_v(s)$ are thus 1 for *possible* and 0 for *impossible*.

In the possibilistic framework, extreme forms of partial knowledge can be captured, namely:

- Complete knowledge: for some state $s_0, \pi_v(s_0) = 1$ and $\pi_v(s) = 0$ for other states s (only s_0 is possible)
- Complete ignorance: $\pi_v(s) = 1, \forall s \in S$, (all states are totally possible).

Two pieces of incomplete information can be compared in terms of information content: a piece of information $v \in A_1$ is said to be *more specific* than a piece of information $v \in A_2$ if and only if A_1 is a proper subset of A_2. In terms of respective possibility distributions, say π_1 for $v \in A_1$ and π_2 for $v \in A_2$ it corresponds to the inequality $\pi_1 < \pi_2$. Note that a possibility distribution always contains some subjectivity in the sense that it represents information possessed by an agent at a given time point, i.e. it reflects an epistemic state. This information is likely to evolve upon the arrival of new pieces of information, in particular it often becomes more specific. The acquisition of a new piece of information comes down to deleting possible values of v. If $v \in A_1$ is more specific than $v \in A_2$, the first epistemic state is accessible from the second one by the acquisition of new information of the same type.

Given a collection of pieces of incomplete information of the form $\{v \in A_i : i = 1, \ldots n\}$ the least arbitrary possibility distribution that represents this collection is the least specific disjunctive set among those that are compatible with each piece of information $v \in A_i$, i.e., $v \in \cap_{i=1,\ldots n} A_i$. It corresponds to computing the possibility distribution $\pi_v = \min_{i=1,\ldots n} \pi_i$. These notions lie at the roots of possibility theory [133, 43], in its Boolean version.

This type of representation of incomplete information can be found in two areas : classical logic and interval analysis. In both settings, either logic or interval analysis, the kind of information represented is the same. What differs is the type of variable used to describe the state space S: Boolean in the first case, numerical in the second one.

In propositional logic, a collection of information items is a set K, often called belief base, of Boolean propositions p_i expressed by well-formed formulas by means of literals and connectives. Given n Boolean variables with domain $\{true, false\}$, then $S = \{true, false\}^n$ is made of 2^n elements called *interpretations*. They are maximal conjunctions of literals, which comes down to assigning a value in $\{true, false\}$ to each variable. Models of K form a disjunctive subset of S containing all interpretations that make all propositions in K true. K is then understood as the conjunction of propositions p_i. If models of p_i form the set A_i, the set of models of K form the set $\cap_{i=1,\ldots n} A_i$, which does correspond to a *possibilistic* handling of incomplete information.

In interval analysis [93], numerical information items take the form of closed real intervals $v_i \in [a_i, b_i]$ describing incomplete knowledge of parameters or inputs of a mathematical model described by a real function f. A typical problem is to compute the set of values of $f(v_1, \ldots, v_n)$ when the v_i's lie in the sets $[a_i, b_i]$ that is, $A = \{f(s_1, \ldots, s_n) : s_i \in [a_i, b_i], i = 1, \ldots n\}$.

2.2 Uncertainty

A piece of information is said to be *uncertain* for an agent when the latter does not know whether this piece of information is true or false. A primitive item of information being a proposition, or the statement that an event occurred or will occur, and such a proposition being modeled by a subset of possible values of the form $v \in A$, one may assign a token of uncertainty to it. This token, or uncertainty qualifier, is located at the metalevel with respect to the pieces of information. It can be numerical or symbolic (e.g. linguistic). For instance, consider the statements :

- the probability that the activity takes more than one hour is 0.7.
- It is very possible that it snows to-morrow.
- It is not absolutely certain that Jean comes to the meeting to-morrow.

In these examples, uncertainty qualifiers are respectively a number (a probability), and symbolic modalities (possible, certain). The most usual representation of uncertainty consists of assigning to each proposition or event A, viewed as a subset of S, a number $g(A)$ in the unit interval. $g(A)$ evaluates the likelihood of A, the confidence of the agent in the truth of proposition asserting $v \in A$. This proposition can only be true or false by convention, even if the agent may ignore this truth value. The following requirements sound natural:

$$g(\emptyset) = 0; \quad g(S) = 1; \tag{1}$$

so is the monotonicity with respect to inclusion:

$$\text{If } A \subseteq B \text{ then } g(A) \leq g(B). \tag{2}$$

Indeed, the contradictory proposition $\emptyset$ is impossible, and the tautology S is certain. Moreover if A is more specific than B in the wide sense (hence implies it), the agent cannot be more confident in A than in B: all things being equal, the more imprecise a proposition, the more certain it is. In an infinite setting, continuity properties with respect to converging monotonic sequences of sets must be added. Under these properties, the function g is is sometimes called a *capacity* (after Choquet [15][1]), sometimes a *fuzzy measure* (after Sugeno [119]). In order to stick to the uncertainty framework, it is here called a *confidence function.* Easy but important consequences of postulates (1) and (2) are:

$$g(A \cap B) \leq \min(g(A), g(B)); \quad g(A \cup B) \geq \max(g(A), g(B)). \tag{3}$$

An important particular case of confidence function is the probability measure $g = P$ which satisfies the additivity property

$$\text{If } A \cap B = \emptyset, \text{ then } P(A \cup B) = P(A) + P(B). \tag{4}$$

[1]originally, with explicit reference to electricity!

Given an elementary piece of incomplete information of the form $v \in E$, held as certain, other types of confidence functions, taking on values in $\{0, 1\}$ can be defined:

- a *possibility measure* Π such that $\Pi(A) = 1$ if $A \cap E \neq \emptyset$, and 0 otherwise
- a *necessity measure* N such that $N(A) = 1$ if $E \subseteq A$, and 0 otherwise.

It is easy to see that $\Pi(A) = 1$ if and only if proposition "$v \in A$" not inconsistent with information item "$v \in E$", and that $N(A) = 1$ if and only if proposition "$v \in A$" is entailed by information item $v \in E$. This is the Boolean version of possibility theory [44].

$\Pi(A) = 0$ means that A is impossible if "$v \in E$" is true. $N(A) = 1$ expresses that A is certain if "$v \in E$" is true. Moreover to say that A is impossible ($A \cap E = \emptyset$) is to say that the opposite event is $\overline{A}$ is certain. So, functions N and Π are totally related to each other by the adjointness property:

$$N(A) = 1 - \Pi(\overline{A}). \tag{5}$$

This adjointness relation is the main difference between necessity and possibility measures on the one hand, and probability measures on the other hand, which are self-adjoint in the sense that $P(A) = 1 - P(\overline{A})$.

Uncertainty of the possibilistic type is clearly at work in classical logic. If K is a base propositional belief base with set of models E, and p is the syntactic form of proposition $v \in A$, then $N(A) = 1$ if and only if K implies p, and $\Pi(A) = 0$ if and only if $K \cup \{p\}$ is logically inconsistent. Note that the presence of p in K means that $N(A) = 1$, while its negation $\neg p$ in K is used to mean $\Pi(A) = 0$. However, in propositional logic, it cannot be expressed that $N(A) = 0$ nor $\Pi(A) = 1$. To do so, a modal logic is needed (Chellas [14]), that prefixes propositions with modalities such as possible ($\Diamond$) and necessary ($\Box$): In a modal belief base K^{mod}, $\Diamond p \in K^{mod}$ encodes $\Pi(A) = 1$, and $\Box p \in K^{mod}$ encodes $N(A) = 1$ (which is encoded by $p \in K$ in classical logic). The adjointness relation (5) is well-known in modal logic, where it reads: $\Diamond p = \neg\Box\neg p$.

It is easy to check that each of possibility and necessity measures saturates one of the inequalities (3):

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)). \tag{6}$$

$$N(A \cap B) = \min(N(A), N(B)). \tag{7}$$

Possibility measures are said to be *maxitive* and characterized (in the finite setting) by the maxitivity property (6). Similarly, necessity measures are said to be *minitive* and are characterized (in the finite setting) by the minitivity property (6). These properties are taken as postulates even when possibility and necessity values lie in $[0, 1]$. In the Boolean setting, they respectively read $\Diamond(p \vee q) = \Diamond p \vee \Diamond q$ and $\Box(p \wedge q) = \Box p \wedge \Box q$ and are well-known in modal logics. In fact, it also hold that $N(A) > 0$ implies $\Pi(A) = 1$, and the Boolean possibilistic setting is thus captured by the modal logic KD45, which is typical of Hintikka's epistemic logic [74].

In general, possibility measures are distinct from necessity measures. Maxitivity and minitivity properties cannot simultaneously hold for all events, except if $N = \Pi$ corresponds to precise information ($E = \{s_0\}$). It then also coincide with a Dirac probability measure, since then $g(A) = 1$ if and only if $g(\overline{A}) = 0$. However note that, it may occur that $N(A \cup B) > \max(N(A), N(B))$ and $\Pi(A \cap B) < \min(\Pi(A), \Pi(B))$. Namely, it is easy to check that if it is not known whether A is true or false (because $A \cap E \neq \emptyset$ and $\overline{A} \cap E \neq \emptyset$), then $\Pi(A) = \Pi(\overline{A}) = 1$ and $N(A) = N(\overline{A}) = 0$; but, by definition $\Pi(A \cap \overline{A}) = \Pi(\emptyset) = 0$ and $N(A \cup \overline{A}) = N(S) = 1$. The possibilistic approach thus distinguishes between three extreme epistemic states:

- The certainty that $v \in A$ is true : $N(A) = 1$, which implies $\Pi(A) = 1$;
- The certainty that $v \in A$ is false : $\Pi(A) = 0$, which implies $N(A) = 0$:
- The ignorance as to whether $v \in A$: $\Pi(A) = 1$, and $N(A) = 0$.

The item of Boolean information $v \in E$ may also lead to define a probability measure. Whenever this is the only available information, the Insufficient Reason principle of Laplace proposes to assign (in the finite setting) the same probability weight to each element in E (by symmetry, i.e. lack of reason not to act so), which comes down to letting

$$P(A) = \frac{Card(A \cap E)}{Card(E)}.$$

The idea is that E should be defined in such a way that all its elements have equal probability. This probability measure is such that $P(A) = 1$ if and only if $E \subseteq A$, and $P(A) = 0$ if and only if $E \cap A = \emptyset$. It plays the same role as the pair (Π, N) and it refines it since it measures to what extent A overlaps E. Nevertheless, probabilities thus computed depend on the number of elements inside E. In the case of total ignorance ($E = S$), some contingent events (different from S and $\emptyset$) will be more probable than others, which sounds paradoxical. The possibilistic framework proposes a less committal representation of ignorance: all contingent events and only them are equally possible and certain (they have possibility 1 and necessity 0). The situation of total ignorance is not faithfully rendered by a single probability distribution.

2.3 Gradual linguistic information

The representation of a proposition as an entity liable to being true or false (or, of an event that may occur or not) is a convention. This convention is not always reasonable. Some kinds of information an agent can assert or understand do not lend themselves easily to this convention. For instance, the proposition *Pierre is young* could be neither totally true, nor totally false: it sounds more true if Pierre is 20 years old than if he is 30 (in the latter case, it nevertheless makes little sense to say that Pierre is not young). Moreover the meaning of *young* will be altered by linguistic hedges expressing intensity : it makes sense to say *very young, not too young*, etc. In other words, the proposition *Pierre is young* is clearly not Boolean. It underlies a ranking, in terms of relevance, of attribute values to which it refers. This kind of information is taken into account by the concept of *fuzzy set* (Zadeh [129]). A fuzzy set F is an application from S to a (usually) totally ordered scale L often chosen as the interval $[0, 1]$. $F(s)$ is the membership degree of element s to F. It is a measure of the adequacy between situation s and proposition F.

It is natural to use fuzzy sets when dealing with a piece of information expressed in natural language and referring to a numerical attribute. Zadeh [130, 131, 132] introduced the notion of *linguistic* *variable* ranging in a finite ordered set of linguistic terms. Each term represents a subset of a numerical scale associated to the attribute and these subsets forms a partition of this scale. For instance, the set of terms $F \in \{young,\ adult,\ old\}$ forms the domain of the linguistic variable *age(Pierre)*. It partitions the scale of this attribute. Nevertheless, if seems that transitions between age zones corresponding to terms are gradual rather than abrupt. In the case of the predicate *young*, it sounds somewhat arbitrary to define a precise threshold $s_\star$ on a continuous scale such that $F(s) = 0$ if $s > s_\star$ and 1 otherwise. Such linguistic terms are so-called *gradual predicates*. You can spot them by the possibility to alter their meaning by intensity adverbs such as the linguistic hedge *very*. The membership scale [0, 1] is but the mirror image of the continuous scale of the attribute (here: the age). Not all predicates are gradual: for instance, it is clear that *single* is Boolean.

It is important to tell degrees of adequacy (often called *degrees of truth*) from degrees of confidence, or belief. Already, within natural language, sentences *Pierre is very young* and *Pierre is probably young* convey different meanings. According to the former sentence, the membership degree of *age(Pierre)* to $F = young$ is clearly high; according to the latter, it is not totally excluded that Pierre is old. A membership degree is interpreted as a degree of adequacy if the value $age(Pierre) = s$ is known and the issue under concern is to provide a linguistic qualifier to describe Pierre. The term *young* is adequate to degree $F(s)$.

The standpoint of fuzzy set theory is to consider any evaluation function as a set. For instance, a utility function can be viewed as a fuzzy set of *good* decisions. This theory defines gradual, non-Boolean extensions of classical logic and its connectives (disjunction, conjunction, negation, implication). Of course, natural questions may be raised such as the measurement of membership functions, the commensurability between membership functions pertaining to different attributes, etc. These are the same questions raised in multifactorial evaluation. Actually, the membership degree $F(s)$ can be seen as a degree of similarity between the value s and the closest prototype of F, namely some s_0 such that $F(s_0) = 1$; $F(s)$ is inversely proportional to the distance between this prototype s_0 and the value s. The membership degree often has such a metric interpretation, which relies on the existence of a distance in S.

When the only available information is of the form $v \in F$, where F is a fuzzy set (for instance *Pierre is very young*) then, like in the Boolean case, the membership function is interpreted as a possibility distribution attached to v: $\pi_v = F$ (Zadeh [133]). But now, it is a gradual possibility distribution on the scale L, here $[0, 1]$. Values s such that $\pi_v(s) = 1$ are the most plausible ones for v. The plausibility of a value s for v is then all the greater as s is close to a totally plausible value.

Possibility theory is driven by the principle of minimal specificity. It states that any hypothesis not known to be impossible cannot be ruled out. A possibility distribution is said to be at least as specific as another one if and only if each state is at least as possible according to the latter as to the former (Yager [126]). In the absence of sufficient information each state is allocated the maximal degree of possibility: this is the minimal specificity principle. Then, the least specific distribution is the least restrictive and informative, or yet the least committal.

Plausibility and certainty evaluations induced by the information item $v \in F$ concerning a proposition $v \in A$ can be computed in terms of possibility and necessity degrees of event A:

$$\Pi(A) = \max_{s \in A} \pi_v(s); \quad N(A) = 1 - \Pi(\overline{A}) = \min_{s \notin A} 1 - \pi_v(s) \tag{8}$$

It is clear that a gradual information item is often more informative than Boolean information: $v \in F$, where F is gradual, is more specific than $v \in A$ when $A = \text{support}(F) = \{s, F(s) > 0\}$, because the former suggests a plausibility ranking between possible values of v in A. This representation of uncertainty through the use of gradual linguistic terms leads to quantifying plausibility in terms of distance to an ideally plausible situation, not in terms of frequency of occurrence, for instance.

2.4 Granularity

In the previous subsections, assumptions that underlie the definition of the set S of states of affairs were not laid bare. Nevertheless the choice of S has a clear impact on the possibility or not to represent relevant information. In Decision Theory, for instance, it is often supposed that S is infinite or detailed enough to completely describe the problem under concern. Nevertheless this assumption is sometimes hard to sustain. On the real line, for instance, only so-called measurable sets can be assigned a probability even if intuitively it should be possible to do so to any event that makes sense in a situation [70]. In fact, using reals numbers is often due to the continuous approximation of information that is intrinsically discrete, or perceived as such. For instance, probability distributions derived from statistical data can be viewed as idealizations of histograms, which are finite entities, not only because representing a finite number of observations, but also from our inability to perceive the difference between very close values. This indistinguishability can be encountered as well when representing preferences of an agent accounting for indifference thresholds on the utility function.

Moreover, the set S seldom comes out of the blue. In the approach by De Finetti [24], just like in the logical approach to Artificial Intelligence, the primitive information consists of a collection of propositions expressed in some prescribed language, to which an agent assigns degrees of confidence. The state space S is then generated from these propositions (mathematically, its subsets form the smallest Boolean algebra containing the subsets of models of these propositions). This way of proceeding has non-trivial consequences for representing and revising information. For instance, if a new proposition is added, it may result in a modification, especially a refinement of the set S. This is called a *granularity change* for the representation. A set S_2 is a refinement of S_1 [104] if there is an onto mapping ρ from S_2 to S_1 such that the reciprocal images of elements in S_1 via ρ, namely the sets $\{\rho^{-1}(s) : s \in S_1\}$ form a partition of S_2 (Zadeh [134] speaks of S_1 being a "granulation" of S_2). It is clear that the probabilistic representation of incomplete information by means of the Insufficient Reason principle does not resist to a change of granularity: the image on S_1 of a uniform probability on S_2 via ρ is not a uniform probability, generally. And this principle applied to S_1 inside equivalence classes of S_2 may not produce a uniform probability on S_2 either. So, the probabilistic representation of ignorance sounds paradoxical as it seems to produce information out of the blue while changing the frame. This anomaly does not appear with the possibilistic representation: the image of a uniform possibility distribution on S_2 via ρ is a uniform possibility distribution indeed. Conversely, applying the minimal specificity (or symmetry) principle in two steps (to S_2, then S_1) produces a uniform possibility distribution on S_1.

The simplest case of granularity change is the following: let Ω be a set of entities described by means of attributes $V_1, V_2, \ldots, V_k$ with respective domains $D_1, D_2, \ldots, D_k$. Then S is the Cartesian product $D_1 \times D_2 \times \cdots \times D_k$. Each element in S can be refined into several elements if a $k+1$th attribute is added. Suppose a collection of individuals Ω described by such attributes. Nothing forbids different individuals from sharing the same description in terms of these attributes. Then let Ξ be a subset of Ω. Generally it is not possible to describe it by means of S. Indeed, let R be the equivalence relation on Ω defined by the identity of descriptions of elements ω of Ω: $\omega_1 R \omega_2$ if and only if $V_i(\omega_1) = V_i(\omega_2), \forall i = 1, \ldots, k$. Let $[\omega]_R$ be the equivalence class of ω. Each element in S corresponds to an equivalence class in Ω. Then, the the set Ξ can only be approximated by the language of S but not exactly described by it. Let Ξ^* and Ξ_* be the upper and lower approximations of Ξ defined as follows

$$\Xi^* = \{\omega \in \Omega : [\omega]_R \cap \Xi \neq \emptyset\}; \quad \Xi_* = \{\omega \in \Omega : [\omega]_R \subseteq \Xi\} \tag{9}$$

The pair (Ξ^*, Ξ_*) is called *rough set* by Pawlak [97]. Only sets Ξ^* and Ξ_* of individuals can be perfectly described by combinations of attribute values $V_1, V_2, \ldots, V_k$ corresponding to the subsets of S. Note that histograms and numerical images correspond to this very notion of indistinguishability and granularity, equivalence classes respectively corresponding to boxes of the histogram and to pixels.

When changing granularity by adding a new attribute that is logically independent from others, each element in S_1 is refined into as many elements in S_2 and a uniform probability on one set is compatible with a uniform probability on the other one. In the case of adding a proposition that is not logically independent from others) the induced refinement is not always that homogeneous.

3 Probability theory

Probability theory is the oldest among uncertainty theories and the best mathematically developed as well as the most widely acknowledged. Probability theory can be envisaged from a purely mathematical side, as often the case since the emergence of Kolmogorov axioms in the 1930's [2]. Under this view, the starting point is a sample space Ω, an algebra of measurable subsets $\mathcal{B}$ and a measure of probability P i.e., a mapping from $\mathcal{B}$ in $[0, 1]$ such that

$$P(\emptyset) = 0; \quad P(\Omega) = 1; \tag{10}$$

$$\text{si } A \cap B = \emptyset \text{ alors } P(A \cup B) = P(A) + P(B). \tag{11}$$

The triple $(\Omega, \mathcal{B}, \mathcal{P})$ is called a probability space. A random variable is construed as a mapping V from Ω in some representation space S (often the real line). In the simplest case, S is supposed to be a finite set, which prescribes a finite partitioning of Ω according to the procedure described in subsection 2.4. The family of measurable sets $\mathcal{B}$ can be defined as the Boolean algebra induced by this partition. The probability distribution associated to the random variable V is then characterized by an assignment of weights $p_1, p_2, \ldots, p_{card(S)}$, to elements of S (i.e., $p_i = P(V^{-1}(s_i))$, such that

$$\sum_{i=1}^{card(S)} p_i = 1.$$

Beyond a basically consensual mathematical framework (up to discussions on the meaning of zero probabilities and the issue of infinite additivity), significantly diverging views of what a probability degree may mean can be found in the literature (Fine [61]). This section reviews some of these

[2] and typically in France where statistics is a branch of applied mathematics.

controversies, emphasizing the limitations of uncertainty representations relying on the use of a unique probability distribution.

3.1 Frequentists and subjectivists

We consider probability theory as a tool for representing information. For this purpose, probabilities must be given a concrete meaning. Traditionally, there is at least three interpretations of probability degrees. The oldest and simplest is in terms of counting equally possible cases. It comes back to Laplace at the turn of the nineteenth century. For instance, Ω is supposed finite and p_i is proportional to the number of elements in $V^{-1}(s_i)$. The probability of an event is the number of favorable cases, where this event occurs, divided by the total number of possible cases. The validity of this approach relies on (i) Laplace's Insufficient Reason principle stating that equally possible states are equally probable and (ii) the capability of constructing S in such a way that its elements are indeed equipossible. This can be helped by appealing to symmetry considerations, justifying assumptions of purely random phenomena (like in games with unbiased coins, dice, etc.).

To-date, the most usual interpretation of probability is frequentist. Observations (that form a relevant sampling of the set Ω), are collected, (say a finite subset $\Omega(n) \subset \Omega$ with n elements). These observations are supposedly independent and made in the same conditions. Frequencies of observing $V = s_i$ can be calculated as :

$$f_i = \frac{card(V^{-1}(s_i) \cap \Omega(n))}{n}$$

or, if S is infinite, a histogram associated to the random variable V can be set up considering frequencies of members of a finite partition of S. It is supposed that, as the number of observations increases, $\Omega(n)$ becomes fully representative of Ω, and that frequencies f_i converge to "true" probability values $p_i = \lim_{n\to\infty} f_i$. The connection between frequency and probability dates back to J. Bernoulli's law of large numbers, proving that when tossing a fair coin a great number of times, the proportion of heads tends to become equal to the proportion of tails.

This definition of probabilities requires a sufficient number of observations (ideally infinite) of the phenomenon under concern. Then, assigning a probability to an event requires a population of situations and reveals a trend in this population. A probability distribution is then viewed as generic knowledge. This framework also forbids to assign probabilities to non-repeatable events. Only statistical prediction is allowed, that is, a degree of confidence in getting "head" in the next toss of the coin reflects the proportion of heads observed so far in a long sequence of experiments. However, the idea that statistical experiments are rigorously repeatable is debatable. The frequentist assumption of independent observations collected in identical conditions is often only approximately verified (one might suspect some contradiction between the identity of experimental conditions and the independence of the observations, like when measuring the same quantity several times with the same sensor). In general, even when they are independent, experimental conditions under which observations are collected may only be similar to one another. A frequentist approach can then still be developed [68]. In the case of non-repeatability, (testimonies, elections, for instance), one is led to a subjectivist view of probabilities, which then directly represent degrees of belief of agent about the occurrence of singular events or the truth of relevant propositions for the problem at hand. This point of view meets a caveat: how to justify the additivity law of probabilities seen as degrees of belief? In the case of repeatable phenomena, considered random, the use of frequencies is in agreement with the additivity axiom (11). What plays the role of frequencies for non-repeatable events, are amounts of money bet on the occurrence or the non-occurrence of singular events (according to suggestions originally made by De Finetti [23] and Ramsey [98].

The degree of confidence of an agent in the occurrence of event A is the price $P(A)$ this agent (call her a player) would accept to pay in order to buy a lottery ticket that brings back 1 euro if event A occurs (and 0 euro if not). The more the player believes in the occurrence of A, the less she finds it risky to buy a lottery ticket for a price close to 1 euro. In order to force the latter to propose a fair price, it is moreover assumed that the person that sells lottery tickets (the banker) will not accept the transaction if prices she finds too low are proposed by the player. In particular if the proposed price is too low, the banker is allowed to exchange roles with the player, in which case the latter is obliged to sell the lottery ticket at price $P(A)$ and to pay 1 euro to the banker if event A occurs. This approach relies on a principle of coherence that presupposes a rational agent, i.e., a player that tries to avoid sure loss. For suppose the player buys two lottery tickets pertaining to two opposite propositions A and $\overline{A}$. The principle of coherence then enforces $P(A) + P(\overline{A}) = 1$. Indeed, only one of the two events A or $\overline{A}$ occurs in this one-shot setting. So prices must be such that $P(A)+P(\overline{A}) \leq 1$, lest the player surely lose $P(A)+P(\overline{A})-1$ euros. But if the player proposes prices such that $P(A) + P(\overline{A}) < 1$ then the banker would turn into a player in order to avoid sure loss too. Similarly, with three mutually exclusive propositions A, B and $\overline{A \cup B}$, it can be shown that only $P(A) + P(B) + P(\overline{A \cup B}) = 1$ is rational, and since $P(\overline{A \cup B}) = 1 - P(A \cup B)$, it follows that $P(A \cup B) = P(A) + P(B)$.

This framework can be used on problems having a true answer, for instance, *what is the birth-date of the current Brazilian president?* Clearly no statistical data can be accurately useful for an agent to answer this question if this agent does not know the answer beforehand. The above procedure might end-up with a subjective probability distribution on possible birth-dates, and the resulting outcome can be checked. Note that here uncertainty is due to incomplete information, while in a coin-tossing experiment, it is due to the variability of the outcomes.

The subjectivist approach sounds like a simple reinterpretation of the frequentist probability framework. Actually, as pointed out by De Finetti [24] and his followers (Coletti and Scozzafava [16]), this is not so straightforward. In the the subjectivist approach, there is no sample space. The starting point is a set of Boolean propositions $\{A_j : j = 1, n\}$ to which an agent assigns coherent degrees of confidence c_i, and a set of logical constraints between these propositions. The state space S is then constructed on the basis of these propositions and constraints. By virtue of the principle of coherence, the agent is supposed to choose degrees of confidence according to some probability measure P in such a way that $c_j = P(A_j), \forall j = 1, \ldots, n$. While the frequentist approach is to start from a unique probability measure (obtained by estimation from statistical data) that models the repeatable phenomenon under study, there is nothing of the like in the subjective setting. There may even be several probability measures such that $c_j = P(A_j), \forall j = 1, \ldots, n$. Each of them is rational, but the available information may not allow to isolate it. There may also be no probability measure satisfying these constraints if the agent is not coherent. Computing the probability $P(A)$ of any event A based on the knowledge of pairs $\{(A_j, c_j) : i = 1, n\}$ requires the solution of a linear programming problem whose variables are probability weights p_i attached to elementary events (De Finetti [23]) namely :

maximise (or minimise) $\sum_{s_i \in A} p_i$ under the following constraints :

$$c_j = \sum_{s_k \in A_j} p_k, \forall j = 1, \ldots, n.$$

In this sense, the subjectivist approach to probability is an extension of the logical approach to knowledge representation, and of classical deduction (see also Adams and Levine [1]). Moreover, the subjectivist approach does not require the σ-additivity of P (i.e., axiom (11) for an infinite denumerable set of mutually exclusive events), contrary to the frequentist Kolmogorovean approach. More differences between subjective and frequentist probabilities can be laid bare when the notion of conditioning comes into play.

3.2 Conditional Probability

It is obvious that assigning a probability to an event is not carried out in the absolute. It is done inside a certain context embodied by the frame S. In practice, S never contains *all* possible states of the world, but only those that our current knowledge or working assumptions do not rule out. For instance, in the dice-tossing problem, S contains the six facets of the dice, not the possibility for the dice to break into pieces. It suggests to write the probability $P(A)$ in the form $P(A \mid S)$ to highlight this aspect. If later on, the agent obtains new pieces of information than lead to restrict the set of states of the world further the context of these probabilities will change. Let $C \subset S$ be the current relevant context, and $P(A \mid C)$ be the probability of A in such a context. The transformation from $P(A)$ to $P(A \mid C)$ essentially consists in renormalizing probabilities assigned to states where C is true, that is:

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} \tag{12}$$

This definition retrieves $P(A)$ under the form $P(A \mid S)$. This definition is easy to justify in the frequentist case, since then, $P(A \mid C)$ is the limit of a relative frequency.

Two known results can then be derived:

- *The total probability theorem* : If $\{C_1, \ldots, C_k\}$ forms a partition of S, then

$$P(A) = \sum_{i=1}^{k} P(A \mid C_i) P(C_i).$$

- *Bayes theorem*:

$$P(C_j \mid A) = \frac{P(A \mid C_j) P(C_j)}{\sum_{i=1}^{k} P(A \mid C_i) P(C_i)}.$$

The first result enables the probability of an event to be computed for a general context S given known probabilities of this event in more specific contexts, provided that these contexts form a partition of possible states, and that probabilities of each of these contexts are known. It is instrumental for backward calculations in causal event trees.

Bayes theorem can deal with the following classification problem: Consider k classes C_j of objects forming a partition of S. If the probability $P(A \mid C_j)$ of property A for objects of each class C_j is known, as well as the prior probabilities $P(C_j), j = 1 \dots, k$ that an object is of class C_j, then for any new object which is known to possess property A, it is possible to derive the probability $P(C_j \mid A)$ that this object belongs to class C_j. In diagnosis problems, replace class by fault type and property by symptom.

Bayes theorem is also instrumental in model inference, or learning from statistical data. Then

- the set of classes is replaced by the range of values $\theta \in \Theta$ of the model parameter,
- $P(A \mid \theta)$ is the likelihood function known when the type of statistical model is known, and θ is fixed,
- the set A represents a series of observed outcomes,
- a prior probability distribution is given on the parameter space Θ (in case of ignorance, a so-called non-informative prior according to the objective Bayesian school is used)
- the posterior probability $P(\theta \mid A)$ is viewed as the new knowledge about the parameter model after observing A, which leads to a possible update of this model.

In a subjectivist framework, the situation of conditioning is different. The probability $P(A \mid C)$ is now assigned by the agent to the hypothetical occurrence of the conditional event $A \mid C$. Conditional probability is now considered as a primitive notion (no longer derived from a probability measure). Namely, $A \mid C$ represents the occurrence of event A in the hypothetical context where C is true. The quantity $P(A \mid C)$ is then still interpreted as an amount of money bet on A, but now this amount is given back to the player if event C does not occur - the bet is then called off (De Finetti [24]). In this operational framework, it can be shown that coherence requires that the equality $P(A \cap C) = P(A \mid C) \cdot P(C)$ be satisfied.

The definition of conditional probability under the form of a quotient presupposes that $P(C) \neq 0$, which may turn out to be too restrictive. Indeed, in the framework proposed by De Finetti, where elicited probabilities may be assigned to any conditional event, the available set of beliefs to be reasoned from takes the form of a collection of conditional probabilities $\{P(A_i \mid C_j), i = 1 \dots m; j = 1 \dots n\}$ corresponding to various potential contexts some of which have zero probability of occurring in the current world. However, by defining conditional probability as any solution to equation $P(A \cap C) = P(A \mid C) \cdot P(C)$, it still makes sense as a non-negative number when $P(C) = 0$ (see details in Coletti and Scozzafava [16]). Besides, in the tradition of probability theory, an event of zero probability is understood as practically impossible, not intrinsically impossible: in other words, it is an exceptional event only (like the dice breaking into pieces). The general reasoning problem in the conditional setting is to compute probability $P(A \mid C)$ from a set of known conditional probabilities $\{P(A_i \mid C_j), i = 1 \dots m; j = 1 \dots n\}$ (Paris [96]), a problem much more general that the one underlying the theorem of total probability.

Under this view, probabilistic knowledge consists of all values $P(A_i \mid C_j)$ known in all contexts. An agent only selects the appropriate conditional probability based on the available knowledge on the current situation, a view in full contrast with the one of revising a probability measure based on the arrival of new knowledge. Indeed, some scholars justify conditional probability as the result of a revision process. The quantity $P(A \mid C)$ is then viewed as the *new* probability of A when the agent hears that event C occurred [65]. Basic to belief revision is the principle of minimal change: the agent minimally revises her beliefs so as to absorb the new information item interpreted by the constraint $P(C) = 1$. A simple encoding of the principle of minimal change is to suppose that probabilities of states that remain possible do not change in relative value, which enforces the usual definition of conditioning (Teller [120]). Another more general approach is to look for the new probability measure P_+ that minimizes an informational distance to the prior probability P under the constraint $P_+(C) = 1$ (Domotor, [30]). If relative entropy is chosen as a measure of distance, it can be shown that P_+ is indeed the conditional probability relative to C. Note that interpreting the context C as the constraint $P_+(C) = 1$ is questionable in the frequentist setting, for, in this case, a probability measure refers to a class of situations (a population), while the information item C often refers to a unique situation (the one of the specific problem the agent tries to solve). Indeed, the constraint $P(C) = 1$ might misleadingly suggest that C is true for the whole population while C occurred only in the specific situation the agent is interested in. In the subjectivist scope, conditioning is but hypothetical, and the known occurrence of C only helps selecting the right reference class.

3.3 The unique probability assumption in the subjective setting

The so-called *Bayesian* approach to subjective probability postulates the unicity of the probability measure that represents beliefs of an agent, as a prerequisite to any further consideration (for instance Lindley [88]). Indeed, if the agent decides to directly assign subjective probabilities to elements of S, the principle of coherence leads to the specification of a unique probability distribution by fear of a sure loss of money (this is also called the Dutch book argument). If the available knowledge is insufficient to uniquely characterize a probability distribution, the Bayesian approach may appeal to selection principles such as the one of Insufficient Reason that exploits the symmetries of a problem, or yet the maximum entropy principle (Jaynes [76], Paris [96]). Resorting to the latter in the subjectivist framework is questionable because it only selects the uniform distribution whenever possible as in the following example

> **Example** Suppose the agent describes her knowledge about a biased coin by providing rough estimates of the probability p of getting a tail. If she considers the bias is towards tail, and if cautious she just provides an estimate p in the form of an interval such as $[0.5, 0.8]$. Applying the maximum entropy principle enforces the choice of the uniform distribution, while selecting $p = 0.65$ (the mid-point of the interval) sounds more sensible and faithful to the trend expressed by the incomplete information supplied by the agent.

In any case (and even in the above example) the Bayesian credo states that any epistemic state of an agent is representable by a unique prior probability distribution. An additional argument in favor of this claim is Savage Decision Theory (see the chapter by Cohen and Jaffray in this volume). It demonstrates that, in an infinite setting, if the agent makes decisions in an uncertain environment, while respecting suitable rationality axioms, in particular the fact that the preference between two acts does not depend on states in which they have the same consequences, then the decision process can be explained as if the agent's knowledge were encoded as a unique probability distribution and decisions were rank-ordered according to their expected utility. Besides, the subjectivist approach is somewhat convergent with the frequentist approach because it is agreed that if the agent possesses reliable statistical information in the form of frequencies, they should be used to quantify belief in the next forthcoming event.

The systematic use of a unique probability as the universal tool for representing uncertainty nevertheless raises some serious difficulties:

- It makes no difference between uncertainty due to incomplete information about a question under concern and uncertainty due variability in past results observed by the agent. In the dice game, how to interpret in a non-ambiguous way a uniform distribution provided by an agent that describes her epistemic state about the dice? Namely, it may be the case that the agent knows the dice is unbiased and that the limit frequency distribution should be uniform (pure randomness). However, if the agent ignores everything about that particular dice, because she was given no chance to try it, then the uniform distribution is but the result of the symmetry principle (the agent has no reason to bet more money on one facet rather than another), and it just expresses ignorance. What it means is that there is no bijection between the possible epistemic states of the agent (which are clearly different in the above two situations) and probability distributions, even if it is correct to consider that the proposed prices for buying the lottery tickets by the player do result from her epistemic state. It does not make perfect sense to *identify* betting rates to degrees of confidence or belief. This limitation in expressivity is somewhat embarrassing in a dynamical framework where the amount of available information evolves, as shown later on: when a new piece of information is obtained, should the agent modify his bets by means of a revision rule, or revise her epistemic state and propose new betting rates accordingly?

- It was pointed out earlier that the choice of frame S depends on the language used, hence on the source of information. One agent may perceive distinct situations another agent will not discern. If several points of view or several languages are simultaneously used in a problem, there will be several frames $S_1, S_2, \ldots, S_p$ (rightly called "frames of discernment" by Shafer [104]) involved to describe the same quantity V, and compatibility relations between these frames. Namely each subset of S_i may only be represented by a rough subset of S_j (in the sense of subsection 2.4). It may become impossible to represent mutually consistent epistemic states on the various frames of discernment by means of a unique probability distribution on each set S_i. Indeed, a uniform distribution on one set may fail to correspond with a uniform distribution on another. For instance, consider the example of the possibility of extra terrestrial life, due to Shafer [104]:

  > **Example**: Generally, people ignore whether there is life or not. Hence $P_1(Life) = P_1(Nolife) = \frac{1}{2}$ on $S_1 = \{Life, Nolife\}$. However if the agent discerns between animal life ($Alife$), and vegetal life only ($Vlife$), with frame $S_2 = \{Alife, Vlife, Nolife\}$, the ignorant agent is bound to propose $P_2(Alife) = P_2(Vlife) = P_2(Nolife) = \frac{1}{3}$. Since $Life$ is the disjunction of $Vlife$ and $Alife$, distributions P_1 and P_2 are incompatible while they are supposed to stem from the same epistemic state.

  The same phenomenon occurs on the continuous real line when a piece of incomplete information of the form $x \in [a, b]$ is represented by a uniform probability density on $[a, b]$. The latter representation is not scale-invariant. Indeed, consider a continuous increasing function

f. Then, stating $x \in [a, b]$ is equivalent to stating $f(x) \in [f(a), f(b)]$. However, if x has a probability distribution with uniform density, the density of $f(x)$ is generally not uniform. It looks as if ignorance on $[a, b]$ can create information on $[f(a), f(b)]$.

- The usual debate between normative and descriptive representations of information is relevant when dealing with uncertainty. If the Bayesian approach is normatively attractive, it may prove to be a poor model to account for the way agents handle confidence degrees (Kahnemann et al. [77]). More recent experimental studies seem to suggest that a human agent may, in some situations, follow the rules of possibility theory instead [99].

- Finally there is a practical measurement difficulty in the case of subjective probabilities. It can hardly be sustained that the agent is capable of supplying, even via price assessments, infinitely precise probability values. What can be expressed consist of fuzzy probabilities (as surprisingly acknowledged even by Luce and Raiffa [89]). Such probabilities would be more faithfully represented by intervals, if not fuzzy intervals. In some situation, they are only linguistic terms (*very probable, quite improbable*, etc...). One may thus argue that subjective probabilities should be represented in a purely symbolic way, or on the contrary, by fuzzy subsets (as in Subsection 2.3) of $[0, 1]$ (see Zadeh [128], Budescu and Wallstein [7], and De Cooman [20]). Some authors even propose higher-order probabilities (for instance, Marschak [90]), which sounds like recursively solving a problem by creating the same problem one step higher.

Note that these defects essentially affect the Bayesian representation of subjective belief in the case of poor information. They are partially irrelevant in the case of frequentist probabilities based on sufficient experimental data. For instance, the lack of scale-invariance of probability densities is no paradox in the frequentist view. If the collected information in terms of values for $x \in [a, b]$ justifies a uniform distribution, it is unsurprising that the encoding of the same information in terms of values for $f(x)$ may not lead to a uniform distribution. But the frequentist framework has no pretense to express subjective ignorance.

These caveats motivated the development of alternative representations of subjective uncertainty. In some of them, the numerical framework is given up and replaced by ordinal structures that underlie subjectivist numerical representations. In other representations, incompleteness is acknowledged as such and injected into probability theory, yielding various approaches, some being more general than others. In all approaches, possibility theory (qualitative or quantitative, respectively [51]) is retrieved as the simplest non trivial non-probabilistic representation of uncertainty.

4 Incompleteness-tolerant numerical uncertainty theories

It is now clear that representations of belief using subjective probabilities, under the Bayesian approach, tend to confuse uncertainty due to variability and uncertainty due to incompleteness of information, on behalf of the principle of symmetry or indifference. This choice of representation is often motivated by the stress put on the subsequent decision step supposed to justify any attempt at representing uncertainty. However, it is legitimate to look for representations of uncertainty that maintain a difference between variability and incompleteness [60]. For instance in risk analysis, an ambiguous response due to a lack of information does not lead to the same kind of decision as when it is due to uncontrollable, but precisely measured variability. In sub-section 2.1, it was pointed out that incompleteness can be conveniently modeled by means of disjunctive sets, in agreement with interval analysis and classical logic. The introduction of incompleteness in uncertainty representations thus comes down to combine disjunctive sets and probabilities. There are two options:

- consider disjunctive sets of probabilities, assuming the agent is not in a position to single out a probability distribution;
- randomise the disjunctive set-based representation of incompleteness of subsection 2.1.

Representing incompleteness goes along with modal notions of possibility and necessity. The generalized probability frameworks will be based on numerical extensions of such modalities. The first line was studied at length by Peter Walley [121], who relies on the use of upper and lower expectations characterizing closed and convex sets of probabilities. The second option, due to Arthur Dempster [25] and Glenn Shafer [104] was further developed by Philippe Smets [116] [112]. It comes down to randomizing the modal logic of incompleteness, assigning to each event so-called degrees of belief and plausibility. It turns out to be a special case of the former, mathematically, but it is philosophically different. In the first theory, the agent represents subjective knowledge by means of maximal buying prices of gambles. The imprecise probability approach can also be interpreted as performing sensitivity analysis on a probabilistic model, i.e., there exists a true probability distribution but it is ill-known and lies in some subjectively assessed probability family. In the Shafer-Smets approach, the agent uses degrees of belief and plausibility without any reference to some unknown probability. Numerical possibility theory [133, 43, 51, 38] whose axioms were laid bare in subsection 2.2 and used in the representation of linguistic information in subsection 2.3 turns out to be a special case of the two above approaches, now interpreted in terms of imprecise probability. The last subsection is dedicated to this special case.

All numerical representations of incompleteness-tolerant uncertainty have the following common feature: the uncertainty of each event A, subset of S is characterized by two, respectively upper and lower, evaluations, we shall respectively call (adopting a subjectively biased language) degrees of *epistemic possibility* and *certainty*, that will be denoted Ep and Cer. Epistemic possibility refers to a lack of surprise. These two degrees define confidence functions on the frame S (in the sense of section 2.2) such that

$$\forall A \subseteq S, Cer(A) \leq Ep(A). \tag{13}$$

They are supposed adjoint to each other, namely:

$$\forall A \subseteq S, Cer(A) = 1 - Ep(\overline{A}). \tag{14}$$

The first condition (13) postulates that an event must be epistemically possible prior to being certain, and the second condition (14) says that an event is all the more certain as its opposite is less epistemically possible. So these functions formally respectively generalize possibility measures ($Ep(A) = \Pi(A)$) and necessity measures ($Cer(A) = N(A)$) of sections 2.2 and 2.3, and probability measures as well($P(A) = Ep(A) = Cer(A)$). This framework has the merit of unambiguously encoding three epistemic states pertaining to event A:

- The case when A is certainly true: $Cer(A) = 1$ (hence $Ep(A) = 1, Ep(\overline{A}) = 0, Cer(\overline{A}) = 0$).
- The case when A is certainly false: $Ep(A) = 0$ (hence $Cer(A) = 0$).
- The case when the agent does not know if A is true or false: $Cer(A) = 0$ et $Ep(A) = 1$ (then $Ep(\overline{A}) = 1; Cer(\overline{A}) = 0$).

The amount of incompleteness of the information pertaining to A is the difference $Ep(A) - Cer(A)$. When information on A is totally missing, there is a maximal gap between certainty and epistemic possibility. The non-certainty of A ($Cer(A) = 0$) is carefully distinguished from the certainty of its negation $\overline{A}$. The distinction between ignorance and what could be understood either as random variability of A (or totally conficting information about it) is also made (the latter is when $Cer(A) = Ep(A) = \frac{1}{2} = P(A)$). The two approaches to the representation of uncertainty presented hereafter, namely, imprecise probabilities and belief functions, do use pairs of set-functions of the (Cer, Ep) kind.

4.1 Imprecise Probabilities

Suppose that the information possessed by an agent is represented by a family of probability measures on S. This approach may sometimes correspond to the idea of imprecise probabilistic model. This imprecision may have various origins:

- In the frequentist framework, the assumptions that frequencies converge may no longer be assumed. At the limit, you only know that the frequency of each elementary event belongs to an interval (Walley and Fine [122]).
- There may be incomplete information about which is the right stochastic model of a repeatable phenomenon. For instance, the nature of a parametric model is known but the value of some parameter like the mean or the variance is incompletely known. Bayesians then choose a prior probability distribution on possible values of parameters. This is precisely what is not assumed by imprecise probabilists [4].
- Pieces of incomplete information are supplied by an agent about a probability distribution (support, mean value, mode, median, some quantiles) in a non-parametric framework.
- In the subjectivist framework, conditional propositions along with (bounds of) probabilities incompletely characterize a subjective probability, after De Finetti [23] and his followers [16].
- Walley [121] gives up the idea of exchangeable bets and allows the agent to propose maximal buying prices and minimal selling prices for gambles that may differ from each other. Gambles are functions from S to the real line, where $f(s)$ is the relative gain in state s, generalizing events. The maximal buying (resp. minimal selling) price of a gamble is interpreted as a lower (resp. upper) expectation, thus defining closed convex sets of probabilities called *credal sets* that can be interpreted as epistemic states [85].
- Gilboa and Schmeidler [66], by relaxing in a suitable way Savage axiom, provide a decision-theoretic justification of the assumption that an agent uses a family of prior probabilities

for making choices among acts. In order to hedge against uncertainty, the agent selects, when evaluating the potential worth of each act, the probability measure ensuring the least expected utility value. See the chapter by Chateauneuf and Cohen in this volume.

In this section, the certainty function $Cer(A)$ and epistemic possibility function $Ep(A)$ are respectively interpreted as lower and upper bounds of a probability $P(A)$ for each event A. The additivity of P forces the following inequalities to be respected by these bounds (Good [70]): $\forall A, B \subseteq S$, such that $A \cap B = \emptyset$,

$$Cer(A) + Cer(B) \leq Cer(A \cup B) \leq Cer(A) + Ep(B) \leq Ep(A \cup B) \leq Ep(A) + Ep(B). \quad (15)$$

Then Cer and Ep are clearly monotonic under inclusion and adjoint to each other (since $Cer(\overline{A})$ must be the lower bound $1 - P(A)$, it follows that $P(A) \geq Cer(A), \forall A$ is equivalent to $P(A) \leq Ep(A), \forall A$). Nevertheless, this approach is not satisfactory as it may be the case that the set of probabilities that function Cer is supposed to bound from below (or, for function Ep, from above), namely the set $\{P : \forall A \subseteq S, P(A) \geq Cer(A)\}$ is empty.

Conversely, we may start from a family $\mathcal{P}$ of probability measures and compute the bounds (Smith [117])

$$P_*(A) = \inf_{P \in \mathcal{P}} P(A); \quad P^*(A) = \sup_{P \in \mathcal{P}} P(A). \quad (16)$$

Letting $Cer(A) = P_*(A)$, and $Ep(A) = P^*(A)$, functions P_* and P^* duly verify properties (13, 14), and (15) as well. P_* and P^* are respectively called *lower and upper envelopes* (Walley [121]). The width of interval $[P_*(A), P^*(A)]$ represents in some way the degree of ignorance of the agent relative to proposition A. When this interval coincides with the whole unit interval, the agent has no information about A. When this interval narrows down to a point, probabilistic information is maximal.

Generally, the only knowledge of upper and lower envelopes of events is not enough to recover $\mathcal{P}$. This is typically the case if $\mathcal{P}$ is not convex. Indeed, the set of probability measures $\mathcal{P}(P_*) = \{P : \forall A \subseteq S, P(A) \geq P_*(A)\}$, called the *core* of P_*, and derived from the lower envelope, is convex (if $P_1 \in \mathcal{P}(P_*)$ and $P_2 \in \mathcal{P}(P_*)$ then, $\forall \lambda \in [0, 1], \lambda \cdot P_1 + (1 - \lambda) \cdot P_2 \in \mathcal{P}(P_*)$) and it contains the convex closure of the original set $\mathcal{P}$. $\mathcal{P}$ and $\mathcal{P}(P_*)$ induce the same lower and upper envelopes. In fact, the strict inclusion $\mathcal{P} \subset \mathcal{P}(P_*)$ may hold even if $\mathcal{P}$ is convex, because upper and lower probability bounds on events cannot characterize the sets of closed convex sets of probability functions. To achieve this characterization, we need all lower expectations of all gambles associated to a convex set $\mathcal{P}$ and the notion of coherence ensuring estimates of these lower expectations are maximal. This is why Walley [121] uses gambles, as generalizations of events, for developing his theory; the logic of gambles is the proper language for describing (convex) credal sets.

Coherent lower probabilities $\underline{P}$ are lower probabilities that coincide with the lower envelopes of their core, i.e. for all events A of X, $\underline{P}(A) = \min_{P \in \mathcal{P}(\underline{P})} P(A)$. It also means for every event A, the bounds are reachable, i.e., there is a probability distribution P in $\mathcal{P}(\underline{P})$ such that $P(A) = \underline{P}(A)$. A characteristic property of a coherent upper probability (hence generated by a non-empty set of probabilities) was found by Giles [69]. Let us use the same notation for A and its characteristic function (a gamble with values in $\{0, 1\}$: $A(s) = 1$ if $s \in A$ and 0 otherwise). A set function Ep is a coherent lower probability if and only if for any family $A_0, A_1, \ldots, A_k$ of subsets of S, and any pair of integers (r, s) such that $\sum_{i=1}^{k} A_i(\cdot) \geq r + s \cdot A_0(\cdot)$, it holds

$$\sum_{i=1}^{k} Ep(A_i) \geq r + s \cdot Ep(A_0).$$

This condition makes sense in terms of gambles and involves optimal minimal selling prices of an agent who sells $k+1$ lottery tickets corresponding to events $A_0, A_1, \ldots, A_k$ and is protected against a sure loss of money. It also provides a tool to compute least upper probability bounds (in case assigned bounds are not optimal), and in this sense, restoring coherence is like achieving a deductive closure in the logical sense. Since all representations considered in this paper correspond to particular instances of coherent lower probabilities, we will restrict ourselves to such lower probabilities on events.

An important particular case of coherence is obtained by weakening probabilistic additivity by a condition stronger than (15), called *2-monotonicity* (Choquet [15]):

$$Cer(A) + Cer(B) \leq Cer(A \cup B) + Cer(A \cap B), \forall A \subseteq S. \tag{17}$$

A 2-monotonic function is also called a *convex capacity.* Its adjoint function Ep is said to be 2-alternating, which corresponds to the property (17) where the inequality is reversed. Due to (17), it is sure that the core $\mathcal{P}(Cer) = \{P : \forall A \subseteq S, P(A) \geq Cer(A)\}$ is not empty and that Cer is a coherent lower probability. However, a coherent lower probability is not always 2-monotone. The property of 2-monotonicity can be extended to k-monotonicity for $k = 3, 4, \ldots$, changing the equality, appearing in the probabilistic additivitity property written with k events, into inequality. However, while probabilistic 2-additivity implies k-additivity $\forall k > 2$, this is no longer true for k-monotonicity: the latter does not imply $k+1$-monotonicity (even if $k+1$-monotonicity does imply k-monotonicity). So there is a countable hierarchy of types of coherent upper and lower probabilities (see Chateauneuf and Jaffray [13]).

An important example of credal set is generated by so-called *probability intervals.* They are defined over a finite space S as lower and upper probability bounds restricted to singletons s_i [19]. They can be seen as a set of intervals $L = \{[l_i, u_i], i = 1, \ldots, n\}$ defining the family

$$\mathcal{P}_L = \{P | l_i \leq p(s_i) \leq u_i, \forall s_i \in S\}.$$

It is easy to see that $\mathcal{P}_L$ is totally determined by only $2|S|$ 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 s_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

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

De Campos et al. [19] have shown that these bounds are coherent and the lower bounds are 2-monotonic.

Another practical example of credal set is a *p-box* [59]. It is defined by a pair of cumulative distributions $(\underline{F}, \overline{F})$ on the real line such that $\underline{F} \leq \overline{F}$, bounding the cumulative distribution of an imprecisely known probability function P. It is a form of generalized interval. The probability family $\mathcal{P}_{p-box} = \{P, \overline{F}(x) \geq P((-\infty, x]) \geq \underline{F}(x), \forall x \in \mathbb{R}\}$ is a credal set. 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.

4.2 Random disjunctive sets and belief functions

The approach adopted in the theory of evidence (Shafer [104]) is somewhat reversed with respect to the one of the imprecise probability schools. Instead of augmenting the probabilistic approach with higher order uncertainty due to incompleteness, described by sets of probabilities, the idea is to inject higher order probabilistic information to the disjunctive set approach to incompleteness. So, instead of a representation of the form $x \in A$ where A is a set of possible values of x, a (generally) discrete probability distribution is defined over the various possible assertions of the form $x \in A$ (assuming a finite frame S). Let m be a probability distribution over the power set 2^S of S. Function m is called *mass assignment*, $m(A)$ the *belief mass* allocated to the set A, and *focal set* any subset A of S such that $m(A) > 0$. Let $\mathcal{F}$ be the collection of focal sets. Usually, no positive mass is assigned to the empty set ($m(\emptyset) = 0$ is assumed). However, the Transferable Belief Model (TBM) after Smets [116] does not make this assumption. Then $m(\emptyset)$ represents the degree of internal contradiction of the mass assignment. The condition $m(\emptyset) = 0$ is a form of normalisation. As m is a probability distribution, the condition $\sum_{A \subseteq S} m(A) = 1$ must hold anyway.

In this hybrid representation of uncertainty, it is important to understand the meaning of the mass function, and it is essential not to confuse $m(A)$ with the probability of occurrence of event A. Shafer [104] says $m(A)$ is the belief mass assigned to A only and to none of its subsets. One may also see $m(A)$ as the amount of probability pending over elements of A without being assigned yet, by lack of knowledge. An explanation in the subjective line consists in saying that $m(A)$ is the probability that the agent only knows that $x \in A$. So, there is an epistemic modality implicitly present in $m(A)$, but absent from $P(A)$. It explains why function m is not required to be inclusion-monotonic. It is allowed to have $m(A) > m(B) > 0$ even if $A \subset B$, when the agent is sure enough that what is known is of the form $x \in A$. In the language of modal logic, one should write $m(A) = P(\Box A)$ where $\Box$ represents a modality such as *the agent only knows that* .... In particular, $m(S)$ is the probability that the agent is completely ignorant.

In practice, a mass assignment results from a situation where the available pieces of information only partially determine the quantity of interest. This is typically the case when only a compatibility relation (instead of a mapping) between a probability space and the frame S of interest to the agent. Let Ω be a set of possible observations and P a probability measure on Ω supposedly available. Suppose there is a multimapping Γ that defines for each value $\omega \in \Omega$ of the quantity v the set $\Gamma(\omega)$ of possible values of the ill-known quantity x in S. If the agent knows $v = \omega$, she only knows that $x \in \Gamma(\omega)$ and nothing else. From the knowledge of a probability function on Ω, only a mass assignment on S is derived, namely: $\forall A \subseteq S, m(A) = P(\{\omega : \Gamma(\omega) = A\})$ if $\exists \omega \in \Omega, A = \Gamma(\omega)$, and 0 otherwise. This technique for generating a mass assignment from a multiple-valued function was proposed by Dempster [25].

Exemple: Consider an unreliable watch. The failure probability ϵ is known. The set Ω describes the possible states of the watch $U = \{KO, OK\}$. The agent cares for the time it is. So, S is the set of possible time-points. Suppose the watch indicates time t. Then the multimapping Γ is such that $\Gamma(OK) = \{t\}$ (if the watch is in order, it provides the right time), and $\Gamma(KO) = S$ (if the watch does not work properly, the time it is is unknown). The induced mass assignment on S is thus $m(\{t\}) = 1 - \epsilon$ and $m(S) = \epsilon$, which is indeed the probability of not knowing the time it is.

The mass assignment obtained in this example is called a *simple support* because the mass is shared between a single subset A of S and S itself. It is a good model of an unreliable source asserting $x \in A$, that an agent believes is irrelevant with probability ϵ. This value is assigned to S so that $m(A) = 1 - \epsilon$.

The probability space Ω can be considered as a sample space like in the framework of frequentist probabilities. But it is then assumed that observations are imprecise.

Example: Consider an opinion poll pertaining to a French presidential election. The set of candidates is $S = \{a, b, c, d, e\}$. There is a population Ω of n individuals that supply their preferences. But since the opinion poll takes place well before the election, individuals may not have made a final choice, even if they do have an opinion. The opinion of individual i is modeled by the subset $\Gamma(i) \subseteq S$. For instance, a left-wing vote is modeled by $\Gamma(i) = \{a, b\}$; for an individual having no opinion, $\Gamma(i) = S$, etc. In this framework, if individual responses of this form are collected, $m(A)$ is the proportion of opinions of the form $\Gamma(i) = A$.

Another method for constructing Γ can be devised when the frame S is multidimensional $S_1 \times S_2, \times \cdots \times S_k$, and a probability distribution P is available on part of the frame, like $S_1 \times S_2, \times \cdots \times S_i$, and there is a set of constraints relating the various parameters $x_1, x_2, \ldots, x_k$, thus forming a relation R on S. R represents all admissible tuples in S. Let $U = S_1 \times S_2, \times \cdots \times S_i$. Then if $u = (s_1, s_2, \ldots, s_i)$, denote $[u]$ the set of tuples in S starting by u; then $\Gamma(\omega) = R \cap [u]$. The above watch example is of this kind.

A mass assignment m induces two set-functions, respectively a belief function Bel and a plausibility function Pl, defined by:

$$Bel(A) = \sum_{E \subseteq A, E \neq \emptyset} m(E); \quad Pl(A) = \sum_{E \cap A \neq \emptyset} m(E). \tag{18}$$

When $m(\emptyset) = 0$, it is clear that $Bel(S) = Pl(S) = 1, Pl(\emptyset) = Bel(\emptyset) = 0$, and $Bel(A) = 1 - Pl(\overline{A})$ so that these functions are another example of certainty ($Cer = Bel$) and epistemic possibility ($Ep = Pl$). Belief functions Bel are k-monotonic for any positive integer k:

$$Bel(\cup_{i=1,\ldots,k} A_i) \geq \sum_{i=1}^{k} (-1)^{i+1} \sum_{I:|I|=i} Bel(\cap_{j \in I} A_j). \tag{19}$$

Plausibility functions satisfy a similar property, reversing the direction of the above inequality.

Conversely, knowing function Bel, a unique mass assignment m can be recomputed from the equations that define $Bel(A)$ for all subsets of S, considering values $m(E)$ as unknowns. This is Moebius transform. This transform, say $M(g)$, actually applies to any set-function g and in particular to the lower envelope P_* of a probability family. Solving these equations is always possible and yields a unique solution in the form of a set-function $m = M(P_*)$ such that $\sum_{A \subseteq S} m(A) = 1$, that however may not be everywhere positive. Links between the cardinality of subsets with positive mass, and the order of the k-monotonicity of a confidence function are studied by Chateauneuf and Jaffray [13]. The positivity of the Moebius transform of a confidence function is characteristic of belief functions. This property shows that belief functions are a special case of coherent lower envelopes, i.e., that $Bel(A) = \inf\{P(A) : P \in \{P : P \geq Bel\}\}$. Nevertheless, this property is generally not exploited in the setting of belief functions. For instance, the Transferable Belief Model by Smets [116] considers $Bel(A)$ as the degree of belief in A for an agent, not as a lower bound of some ill-known objective or subjective probability. This non-probabilistic point of view affects calculation rules (for conditioning, or combination) that must then be devised independently, instead of being induced by probability theory. Smets [111] tried to justify $Bel(A)$ as a genuine non-probabilistic degree of belief through an axiomatic derivation.

Two important particular cases of belief functions must be pointed out:

- Probability functions are retrieved by assuming focal sets are singletons. It is clear that if $m(A) > 0$ implies $\exists s \in S, A = \{s\}$, then $Bel(A) = Pl(A) = P(A)$ for the probability function such that $P(\{s\}) = m(\{s\}), \forall s \in S$. Conversely, Bel is a probability function if and only if $Bel(A) = Pl(A), \forall A \subseteq S$.
- Plausibility functions are possibility measures (or via adjunction, belief functions are necessity measures) if and only if focal sets are nested, i.e., $\forall A \neq B \in \mathcal{F}, A \subset B$ or $B \subset A$. Then, $Pl(A \cup B) = \max(Pl(A), Pl(B))$ and $Bel(A \cap B) = \min(Bel(A), Bel(B))$.

Belief functions have first been defined on finite frames. Their extension to infinite sets pose tricky mathematical problems in the general case [106]. Nevertheless, it is possible to define a belief function on the reals, based on a continuous mass density bearing on closed intervals [118]. For any pair of real numbers $x \leq y$, the mass density $m([x, y])$ is defined by the bi-dimensional probability density $p(x, y)$ taking value 0 if $x > y$. Then, belief and plausibility degrees of intervals of the form $[-\infty, s]$ (which are actually a lower cumulative distribution $F_*(s) = Bel([-\infty, s])$ and an upper distribution $F^*(s) = Pl([-\infty, s])$, respectively) can be obtained as integrals of $p(x, y)$ on the respective domains $\{(x, y), y \leq s\}$ and $\{(x, y), x \leq s\}$. More details in Smets [114]. Contrary to the case of probabilities, these cumulative functions are not sufficient to reconstruct the mass density function (except when focal intervals are nested), nor to compute belief and plausibility or other events. Clearly the pairs (F_*, F^*) are p-boxes that provide a useful summary of the information contained in a belief function, when the question of interest is one of violating a threshold. The lack of information is all the greater as F_* and F^* stand far away from each other. The credal set $\mathcal{P}(\underline{F}, \overline{F})$ induced by any p-box is in fact representable by a belief function whose focal elements are of the form $\{x, \overline{F}(x) \geq \alpha\} \setminus \{x, \underline{F}(x) \geq \alpha\}$ [82]. However, the belief function equivalent to the probability box induced by a belief function is less informative than the original one.

Smets [110] tried to reconcile the theory of exchangeable bets (justifying subjective probabilities) and the postulate that beliefs of an agent are represented by belief functions. A major objection to subjective probability theory is its lack of distinction between situations of known variability (unbiased dice) and ignorance (unknown dice), as emphasized in Section 3.2. The theory of belief functions enables this distinction to be captured: the case of total ignorance is expressed by the mass assignment $m^?(S) = 1$, encoding a situation where $Bel(A) = 0, Pl(A) = 1, \forall A \neq S, \emptyset$ (corresponding to the uninformative possibility distribution $\pi^?$ in Section 2.1). In contrast, a uniform probability distribution correctly expresses that all realisations of a variable v are known to be equiprobable.

If an agent ignores all about variable v, she is thus led to propose a uniform probability distribution on S, following the Insufficient Reason principle of Laplace; if the agent has some knowledge in the form of a belief function with mass assignment m, Smets [110] suggests that the agent should bet with a probability distribution defined by replacing each focal set E by a uniform probability distribution with support E, then computing the convex mixing of these probabilities, weighted by masses $m(E)$. This is the so-called *pignistic* probability defined by the distribution $BetP$:

$$BetP(s) = \sum_{E:s\in E} \frac{m(E)}{Card(E)} \tag{20}$$

This transformation of a belief function into a probability function was originally proposed by Dubois and Prade [39] with a view to generalize Laplace principle. Smets [110] provided an axiomatic justification, finding the probability function satisfying a linearity property (the pignistic probability of a convex sum of belief functions is the convex sum of their pignistic probabilities) and a property of anonymity (the pignistic probability of an event should not change when realisations of this event are exchanged). It turns out that the pignistic probability has been known in cooperative game theory since the 1950's under the name *Shapley value*. Smets axioms are mathematically the same as the ones proposed by Shapley [107] in a quite different context.

Belief functions can be compared in terms of their informative content. Note that belief functions model at the same time imprecise and uncertain information, and one may wish to evaluate their imprecision and their uncertainty separately. A natural imprecision index of a belief function is the expected cardinality of its mass assignment :

$$Imp(m) = \sum_{E\subseteq S} m(E) \cdot Card(E) \tag{21}$$

It is clear that $Imp(m^?) = Card(S)$ and $Imp(m) = 1$ if the mass assignment is a probability. It can be checked that $Imp(m) = \sum_{s\in S} Pl(\{s\})$, i.e. it only depends on the plausibility of the singletons. This numerical index is in agreement with relations comparing belief functions in terms of their imprecision:

- A mass assignment m_1 is said to be at least as specific as a mass assignment m_2 if $\forall s \in S, Pl_1(\{s\}) \leq Pl_2(\{s\})$. This is a natural requirement due to the property of the cardinality-based imprecision index, viewing the function $Pl(\{s\})\forall s \in S$ (called *contour function* by Shafer [104]) as a possibility distribution.
- A mass assignment m_1 is said to be more precise than a mass assignment m_2 if and only if for all events A, the interval $[Bel_1(A), Pl_1(A)]$ is included in the interval $[Bel_2(A), Pl_2(A)]$. Due to the adjunction property between Pl and Bl, it is enough that inequality $\forall A, Pl_1(A) \leq$

$Pl_2(A)$, holds. In other words, the narrower the interval $[Bel(A), Pl(A)]$, the closer it is to a single probability. If $\mathcal{P}(m) = \{P, P(A) \leq Pl(A), \forall A\}$, m_1 is more precise than m_2 means that the credal set $\mathcal{P}(m_1)$ is a subset of $\mathcal{P}(m_2)$. The function m is thus maximally precise when it coincides with a unique probability, and minimally precise if $m = m^?$.

- A mass function m_1 is a *specialization* of a mass assignment m_2 if and only if the three following conditions are verified:
  - Any focal set of m_2 contains at least one focal set of m_1.
  - Any focal set of m_1 is included in at least one focal set of m_2
  - There is a stochastic matrix W whose term w_{ij} is the fraction of the mass $m_1(E_i)$ of the focal set E_i of m_1 that can be reallocated to the focal set F_j of m_2 so as to retrieve the mass $m_2(F_j)$, namely, $m_2(F_j) = \sum_i w_{ij} \cdot m_1(E_i)$, with constraint $w_{ij} > 0$ only if $E_i \subseteq F_j$.

  The latter relation is more demanding than the former ones : if m_1 is a specialisation of m_2, then m_1 is also more precise and more specific than m_2. It is also obvious that if m_1 is a specialization of m_2, then $Imp(m_1) \leq Imp(m_2)$. The converse properties do not hold. Comparing contour functions is less demanding than comparing plausibilities, and $Pl_1 < Pl_2$ does not imply that m_1 is a specialisation of m_2 (see Dubois and Prade [41]).

Example : $S = \{s_1, s_2, s_3\}$. Suppose $m_1(\{s_1, s_2\}) = \frac{1}{2}$ and $m_1(\{s_1, s_3\}) = \frac{1}{2}$; $m_2(\{s_1\}) = \frac{1}{2}$ and $m_2(S) = \frac{1}{2}$. It is easy to see that none of these mass assignments is a specialization of the other one (the inclusion requirements between focal sets are violated). But m_1 is less precise than m_2 (because $Pl_1(A) = Pl_2(A)$ except if $A = \{s_2, s_3\}$, for which $Pl_1(\{s_2, s_3\}) = 1 > Pl_2(\{s_2, s_3\}) = 0.5$). However the two contour functions are the same.

The uncertainty of a belief function can be evaluated by a generalization of entropy $H(P) = -\sum_{i=1}^{card(S)} p_i \cdot lnp_i$. Several extensions were proposed (Dubois and Prade [42]):

- A measure of dissonance: $D(m) = -\sum_{E \subseteq S} m(E) \cdot lnPl(E)$, maximal for uniform probability distributions, minimal ($= 0$) as soon as $Pl(E) = 1$ for all focal sets E (i.e., they intersect : $\cap\{E : m(E) > 0\} \neq \emptyset$).
- A measure of confusion: $D(m) = -\sum_{E \subseteq S} m(E) \cdot lnBel(E)$, high [3] for uniform mass assignments over all sets with cardinality $\frac{card(S)}{2}$, and minimal ($= 0$) as soon as $m(E) = 1$ for some focal set (incomplete and crisp information).
- Klir and Parviz [79] proposed to measure the uncertainty of a mass assignment m by means of the entropy of its pignistic probability, which does evaluate the amount of indecision of an agent faced with a betting situation under uncertainty. More recently, other suggestions include maximizing and minimizing $H(P)$ when P ranges in the credal set associated with the belief function.

[3] In fact maximizing the index obtained by deleting the logarithm (and the minus sign) from this expression; see Dubois and Ramer [57]

4.3 Quantitative Possibility Theory

Like imprecise probability and evidence theories, possibility theory represents uncertainty by means of two adjoint set function: a necessity measure N that is 'minitive', and a possibility measure Π that is 'maxitive'. They have already been introduced above in sections 2.2 and 2.3. Nevertheless, in this section, one sees these set-functions as lower and upper probabilities, since they can be generated from mass functions associated to nested focal sets. While Zadeh [133] defines possibility distributions from linguistic pieces of information, the idea of considering possibility measures as counterparts to probability measures is due to the economist G.L.S. Shackle [103] who named degree of potential surprise of event A the quantity $N(\overline{A}) = 1 - \Pi(A)$. Possibility theory, in its numerical variant, proposes a very simple model of uncertainty tailored for imprecise information and it can encode particular families of probabilities in a very concise way. This model not only enables us to represent linguistic information (according to Zadeh), but it also generalizes the set-based representation of information (propositional logic, interval analysis), and it can, in an approximate way, represent imprecise statistical information [38].

4.3.1 Possibility theory and belief functions

More precisely, let m be a mass function on a finite set S. One defines the possibility distribution π induced by m, also called its contour function, by letting $\pi(s) = \text{Pl}(\{\text{s}\})$ (plausibility of singletons), i.e.:

$$\forall s \in S, \pi(s) = \sum_{s \in E} m(E). \tag{22}$$

It is easy to see that π takes its values on [0, 1], is normalized ($\pi(s) = 1$ for some state $s \in S$) as soon as the focal sets have a common non-empty intersection (it is in particular the case when they are nested). Recovering m from π is possible only when the focal sets are nested or disjoint. Assume that the focal sets are nested. Then they can be rank-ordered in an increasing sequence $E_1 \subset E_2 \subset, \ldots, \subset E_n$, where $E_i = \{s_1, \ldots s_i\}$, then

$$\pi(s_i) = \sum_{j=i}^{n} m(E_j).$$

The possibility and necessity measures Π and N, defined by the equations (8) from π, coincide with the plausibility and belief functions induced by m. Then the mass function can be recomputed from π as follows (letting $\pi(s_{n+1}) = 0$):

$$m_\pi(E_i) = \pi(s_i) - \pi(s_{i-1}), i = 1, \ldots, n. \tag{23}$$

Thus, we can see that, in the finite consonant case, m_π and π contain the same information, and then $Pl = \Pi$ and $Bel = N$. However, in the infinite case, the relation between consonant random sets and possibility measures is more complex in the general case (see Miranda et al. [91, 92]).

For possibility measures, the precision and specialisation orderings coincide with the specificity ordering of possibility distributions on the singletons: m_{π_1} is a specialisation of m_{π_2} if and only if $\Pi_1(A) \leq \Pi_2(A), \forall A \subseteq S$ if and only if $\pi_1(s) \leq \pi_2(s), \forall s \in S$ (Dubois and Prade [41]).

In the general case, π is only an approximation of m, and it can be checked that π is the least specific possibility distribution such that $Pl \geq \Pi$ and $Bel \leq N$ (Dubois and Prade [46]). It is worth noticing that if the focal sets are imprecise observations coming from a random experiment, equation (22) represents the possibilistic counterpart of an histogram.

4.3.2 Possibility theory and imprecise probabilities

As belief functions mathematically correspond to a particular case of family of probability measures, it is a fortiori the case for possibility distributions. Let us again consider the case of an increasing sequence of nested sets $E_1 \subset E_2 \subset, \ldots, \subset E_k$. Let $\nu_1 \leq \nu_2 \leq, \ldots, \leq \nu_k$, be lower bounds of probability, and let $\mathcal{P}= \{P, P(E_i) \geq \nu_i, \forall i = 1, \ldots, k\}$. This is typically the kind of information provided by an expert who expresses himself in an imprecise way about the value of a parameter. He suggests that $x \in E_i$ with a confidence degree a least equal to ν_i. Then $P_*(A) = \inf_{P \in \mathcal{P}} P(A)$ is a necessity measure; and $P^*(A) = \sup_{P \in \mathcal{P}} P(A)$ is a possibility measure, based on the possibility distribution (Dubois and Prade [48]):

$$\forall s \in S, \pi(s) = \min_{i=1,\ldots,k} \max(E_i(s), 1 - \nu_i). \tag{24}$$

with $E_i(s) = 1$ if $s \in E_i$ and 0 otherwise. See (De Cooman and Aeyels [21]) for an extension of this result to the infinite case. In this framework, each E_i is a kind of confidence set (an interval in the case where $S = \mathbb{R}$) and the probability of belonging to this set is at least ν_i. The probability of not belonging to E_i is thus at most $1 - \nu_i$. This confidence set weighted by a certainty degree corresponds to the possibility distribution $\max(E_i(s), 1 - \nu_i)$. The above equation is nothing but the conjunction of these local distributions. It is clear that distribution π encodes in a very compact way the family of probabilities $\mathcal{P}$. Conversely, a possibility distribution π encodes the credal set defined by $\mathcal{P}(\pi) = \{P, P(A) \leq \Pi(A), \forall A \text{ measurable}\} = \{P, P(A) \geq N(A), \forall A \text{ measurable}\}$.

In the case where $S = \mathbb{R}$, an important particular case of possibility distribution is a *fuzzy interval*. Distribution π is supposed to be upper semi-continuous and quasi-concave ($\forall a, b, c \in \mathbb{R}, \pi(c) \geq \min(\pi(a), \pi(b))$); its level cuts $\{s, \pi(s) \geq \alpha\}, \alpha \in (0, 1]$ are then nested closed intervals $[a^-_\alpha, a^+_\alpha]$. One can associate to π a mass density m_π uniformly distributed over its level cuts. The lower and upper cumulative functions $F_*(s) = N([-\infty, s])$ and $F^*(s) = \Pi([-\infty, s])$ are respectively of the form:

$$F^*(s) = \pi(s), s \in (-\infty, a^-_1], \text{ and 1 if s} \geq a^-_1; \tag{25}$$

$$F_*(s) = 1 - \pi(s), s \in [a^+_1, +\infty) \text{ and 0 if s} \leq a^+_1. \tag{26}$$

Let us consider an interval $A = [x, y]$ including the core of π. The inequality $P(A) \leq \Pi(A)$ writes $F(x) + 1 - F(y) \leq \max(\pi(x), \pi(y))$ where F is the cumulative function of P. One can check that the credal set $\mathcal{P}(\pi)$ is precisely equal to $\{P, \forall x \leq a^-_1, \forall y \geq a^+_1, F(x) + 1 - F(y) \leq \max(\pi(x), \pi(y))\}$. It is generally strictly included in the credal set $\{P, F^* \geq F \geq F_*\}$ [47] of the corresponding p-box. The *mean interval* $[e^-(\pi), e^+(\pi)]$ of π is the set of mean values of the probability distributions in $\mathcal{P}(\pi)$. Its bounds are nevertheless the mean values respectively induced by F^* and F_*.

4.3.3 Clouds and generalized p-boxes

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. It comes down to a family of nested confidence sets $\emptyset \subseteq A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subset S$, with $A_i = \{s_1, \ldots, s_i\}$. Consider two cumulative distributions according to this ordering, that form a p-box. The credal set $\mathcal{P}$ can then be represented by the following restrictions on probability measures

$$\alpha_i \leq P(A_i) \leq \beta_i \qquad i = 1, \ldots, n \tag{27}$$

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 $S = \mathbb{R}$ and $A_i = (-\infty, s_i]$, it is easy to see that we retrieve the usual definition of a p-box.

The credal set $\mathcal{P}$ described by such a generalized P-box can be encoded by a pair of possibility distributions π_1, π_2 s.t. $\mathcal{P} = \mathcal{P}(\pi_1) \cap \mathcal{P}(\pi_2)$ 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 [29].

A *cloud* [95] can be seen as an interval-valued fuzzy set F such that $(0, 1) \subseteq \cup_{x \in S} F(s) \subseteq [0, 1]$, where $F(s)$ is an interval $[\delta(s), \pi(s)]$. It implies that $\pi(s) = 1$ for some s (π is a possibility distribution) and $\delta(s') = 0$ for some s' ($1 - \delta$ is also a possibility distribution). A probability measure P on S is said to belong to a cloud F if and only if $\forall \alpha \in [0, 1]$:

$$P(\delta(s) \geq \alpha) \leq 1 - \alpha \leq P(\pi(s) > \alpha) \tag{28}$$

under all suitable measurability assumptions. From this definition, a cloud (δ, π) is equivalent to the cloud $(1 - \pi, 1 - \delta)$. If S is a finite space of cardinality n, let $A_i = \{s_i, \pi(s_i) > \alpha_{i+1}\}$ and $B_i = \{s_i, \delta(s_i) \geq \alpha_{i+1}\}$. A cloud can thus be defined by the following restrictions [29]:

$$P(B_i) \leq 1 - \alpha_i \leq P(A_i) \text{ and } B_i \subseteq A_i \quad i = 1, \ldots, n \tag{29}$$

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

Let $\mathcal{P}(\delta, \pi)$ be the credal set described by the cloud (δ, π) on a frame S. Clouds are closely related to possibility distributions and p-boxes as follows [29]:

- $\mathcal{P}(\delta, \pi) = \mathcal{P}(\pi) \cap \mathcal{P}(1 - \delta)$ using the credal sets induced by the two possibility distributions π and $1 - \delta$.
- A cloud is a generalized p-box with $\pi_1 = \pi$ and $\pi_2(s_{i+1}) = 1 - \delta(s_i)$ 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 is a generalized p-box and it 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 $1 - \delta$:

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

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 $\pi(s_i)$'s are decreasingly ordered) $\delta(s_i) = \pi(s_{i+1})$ (with $\pi(s_{n+1}) = 0$). Then, $\mathcal{P}(\delta, \pi)$ contains a single probability distribution p such that $p(s_i) = \pi(s_i) - \pi(s_{i+1})$. 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 strong complementarity between possibilistic and probabilistic representations of uncertainty is noticeable. While a probability distribution naturally represents precise pieces of information with their variability (what is called statistical data), a possibility distribution encodes imprecise, but consonant, pieces of information (what is expressed by the nestedness of focal sets). One may consider that the possibilistic representation is more natural for uncertain subjective information, in the sense that from a human agent one rather expects consonant pieces of information, with some imprecision, rather than artificially precise but scattered pieces of information. The fact that a probability measure is lower bounded by a necessity measure and upper bounded by a possibility measure ($N(A) \le P(A) \le \Pi(A), \forall A$) expresses a compatibility principle between possibility and probability: for any event, being probable is more demanding than being possible, and being somewhat certain is more demanding than being probable (Zadeh [133]). A probability measure P and a possibility measure Π are said to be *compatible* if and only if $P \in \mathcal{P}(\pi)$.

4.3.4 Possibility-probability transformations

It is legitimate to look for transformations between probabilistic and possibilistic representations of information. There are several reasons for that. On the one hand, with a view of fusing heterogeneous pieces of information (linguistic pieces of information, measurements issued from sensors), one may wish to have a unique representation framework at our disposal. Besides, the useful information extracted from probability distributions is often much less informative than the original distribution (a prediction interval, a mean value...). Conversely, the subjectivist interpretation of probabilities by the betting theory can be regarded as a probabilistic formalization of the often incomplete pieces of information provided by an agent. Lastly, possibility theory allows us to systematize notions that already exist in the practice of statisticians under an incompletely developed form. The transformation between a probability measure P and a possibility measure Π should obey natural requirements :

- **Possibility-probability consistency** : P and Π should be compatible.
- **Ordinal faithfulness**. One cannot require the equivalence between $P(A) \ge P(B)$ and $\Pi(A) \ge \Pi(B), \forall A, B \subseteq S$, since the ordering induced on the events by P will be always more refined than the one induced by Π. Then one should only insure an ordinal equivalence between the distributions p and π, i.e., $p(s_i) \ge p(s_j)$ if and only if $\pi(s_i) \ge \pi(s_j), \forall s_i, s_j \in S$.

One may also only require a weaker ordinal equivalence, for instance considering that $p(s_i) > p(s_j)$ implies $\pi(s_i) > \pi(s_j)$ but $p(s_i) = p(s_j)$ does not entail $\pi(s_i) = \pi(s_j)$.

- **Informativity**. Probabilistic representation is more precise, thus richer than possibilistic representation. Information is lost when going from the first to the second, information is gained in the converse way. From possibility to probability, one should try to preserve the symmetries existing in the possibilistic representation. From probability $\rightarrow$ to possibility, one should try to lose as little information as possible if the probability measure is statistically meaningful. The case of a subjective probability is different since it often corresponds to poorer knowledge artificially increased by the probabilistic representation, so that a least commitment principle might prevail.

From possibility to probability

For changing a possibility distribution into a probability distribution, it is natural to use the pignistic transformation. If $card(S) = n$, let us denote $\pi_i = \pi(s_i), i = 1, \ldots, n$, assuming that $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_n$. The pignistic transform is a probability distribution p ordinally equivalent to π, such that $p_1 \geq p_2 \geq \cdots \geq p_n$, with $p_i = p(s_i), i = 1, \ldots, n$:

$$p_i = \sum_{j=i}^{n} \frac{\pi_j - \pi_{j+1}}{j}, \forall i = 1, \ldots, n \tag{30}$$

In the case of a fuzzy interval, the mass density associated to $[a^-_\alpha, a^+_\alpha]$ is changed into a uniform probability over this interval, and one considers the uniform probabilistic mixture obtained by integrating over $\alpha \in [0, 1]$. This amounts to build the probability measure of the process obtained by picking a number $\alpha \in [0, 1]$ at random, and then an element $s \in [a^-_\alpha, a^+_\alpha]$ at random (Chanas and Nowakowski [10]). The mean value of the pignistic probability is the middle of the mean interval of π introduced in the previous subsection 4.3.2. This transformation generalizes Laplace Insufficient Reason principle, since applied to a uniform possibility distribution over an interval, it yields the corresponding uniform probability.

From subjective probability to possibility

For the converse change, from probability to possibility, one should distinguish the case where one starts from a subjective probability from the situation where there exist statistics justifying the probability distribution. In the subjectivist framework, and in agreement with the debatable nature of the unique probability provided by an expert, one assumes that the knowledge of the agent is a belief function with mass m over a finite frame S. The elicitation process forces him to provide a probability distribution p that is considered as being the pignistic transform of m. By default, one considers that the least biased belief function is the least informative one, if it exists, among those whose pignistic transform is p (Smets [113]). If one looks for the mass assignment that maximizes the imprecision index $Imp(m) = \sum_{j=i}^{n} \pi_j$ (equation (21)), it can be proved that this mass assignment is unique, that it is consonant, and that it is also minimally specific (w. r. t. the plausibility of singletons) (Dubois et al. [56]). By noticing that the pignistic transformation is a one-to-one mapping between probability and possibility, the least biased representation of the agent's knowledge leading to the subjective probability distribution p is obtained by reversing the equation (30):

$$\pi_i = \sum_{j=1}^{n} \min(p_i, p_j), \forall i = 1, \ldots, n \tag{31}$$

This transformation has been independently introduced by Dubois and Prade [40].

From objective probability to possibility

In the case of an objective probability distribution, one should try, when changing representation, to lose as little information as possible. This leads to look for a possibility distribution π among the most specific ones such that $P \in \mathcal{P}(\pi)$ and which is ordinally equivalent to p.

Let us first consider the discrete case. If $p_1 > p_2 > \cdots > p_n$, and denoting $E_i = \{s_1, \ldots s_i\}$, it is enough to let $\Pi(E_i) \geq P(E_i) \forall i = 1, \ldots, n$ in order to make p and π compatible. By forcing equalities, one gets a unique possibility distribution, maximally specific and ordinally equivalent yo p (Dubois and Prade [39]):

$$\pi_i = \sum_{j=i}^{n} p_j, \forall i = 1, \ldots, n \tag{32}$$

Unicity is preserved when the inequalities between the p_i are no longer strict but the transformation writes $\pi_i = \sum_{j:p_j \leq p_i} p_j, \forall i = 1, \ldots, n$, which maintains the constraint : two equiprobable states are equipossible. If we relax this constraint, one may get possibility distributions compatible with p that are more specific than this one. In particular, equation (32) always yields a possibility distribution that is maximally specific and consistent with p. For instance, if p is a uniform distribution, there are $n!$ ways of ordering S, and equation (32) gives $n!$ non-uniform possibility distributions, maximally specific and consistent with p.

In the case of a unimodal continuous density p over $\mathbb{R}$, this possibility-probability transformation can be extended by considering the level cuts of p, i.e. the subsets $E_\lambda = \{s, p(s) \geq \lambda\}, \lambda \in (0, \sup p]$. If we denote $E_\lambda = [x(\lambda), y(\lambda)]$, then the possibility distribution maximally specific and ordinally equivalent to p is defined by

$$\pi(x(\lambda)) = \pi(y(\lambda)) = 1 - P(E_\lambda) \tag{33}$$

Indeed,it can be proved more generally that, if $P(E_\lambda) = q$, the measurable set A having the smallest measure such that $P(A) = q$ is E_λ (Dubois et al. [53]; Dubois et al. [35]). If p is unimodal, E_λ is the interval with length $L = y(\lambda) - s(\lambda)$ that is the most legitimate representative of the probability density p, in the sense where E_λ is the interval with length L having maximal probability: $P(E_\lambda) \geq P([a, b]), \forall a, b$ such that $b - a = L$. Thus, the transformation (33) can be related to a view of a prediction interval as an imprecise substitute of a probability density, with a given confidence level (often 0.95). Most of the time, this type of interval is defined for symmetrical densities and the considered intervals are centered around the mean. The interval with confidence 0.95 is often defined by the 0.025 and 0.975 percentiles. Characterizing the prediction interval with confidence 0.95 by these percentiles when the distributions are non-symmetrical is not very convincing since this may eliminate values with higher density than the one of some values in this interval. It is much more natural to look for λ such that $P(E_\lambda) = 0.95$. More generally, the α level cut of the possibility distribution π obtained by (33) from p, is the smallest interval with confidence $1 - \alpha$ deducible from p. One can find in the statistical literature a proposal for comparing probability densities according to their 'peakedness' (Birnbaum [5]). It comes down to comparing their possibilistic transforms in terms of their relative specificity; moreover the information ordering of probability measures by means of the entropy index refines the partial specificity of their possibilistic transforms; see Dubois and Huellermeier [31] on this topic.

The transformation (33) builds a family of nested sets around the mode of p. One may systematically build a possibility measure consistent with p by considering any characteristic value s^* in the support of p, and a family of subsets A_λ nested around s^*, indexed by $\lambda \in [0, \omega]$ such that $A_\omega = \{s^*\}$, and $A_0 = support(p)$. For instance, if s^* is the mean mea of p, with standard deviation σ, and that one takes $A_\lambda = [mea - \lambda \cdot \sigma, mea + \lambda \cdot \sigma]$, Chebychev inequality gives us $P(\overline{A_\lambda}) \leq \min(1, \frac{1}{\lambda^2})$. The possibility distribution obtained by letting $\pi(mea - \lambda \cdot \sigma) = \pi(mea + \lambda \cdot \sigma) = \min(1, \frac{1}{\lambda^2})$ is thus consistent with any probability measure with mean mea and standard deviation σ. The probability-possibility transforms can thus yield probabilistic inequalities. It has been shown that a symmetrical triangular possibility distribution with bounded support $[a, b]$ is consistent with any unimodal symmetrical probability function having the same support, and contains the prediction intervals of all these probability measures (Dubois et al. [35]). Moreover, it is the most specific one having these properties (it is consistent with the uniform density over $[a, b]$). This provides, for this distribution family, a probabilistic inequality that is much stronger than the one of Chebychev, and justifies the use of triangular fuzzy intervals for representing incomplete probabilistic information. See Baudrit and Dubois [2] for possibilistic representations of probability families induced by partial knowledge of distribution characteristics.

4.4 Possibility theory and non-Bayesian statistics

Another interpretation of numerical possibility distributions is the likelihood function in non-Bayesian statistics (Smets, [109], Dubois et al. [36]). In the framework of an estimation problem, one is interested in the determination of the value of some parameter $\theta \in \Theta$ that defines a probability distribution $P(\cdot \mid \theta)$ over S. Suppose that we observed event A. The function $P(A \mid \theta), \theta \in \Theta$ is not a probability distribution, but a likelihood function $\mathcal{L}(\theta)$: A value a of θ is considered as being all the more plausible as $P(A \mid a)$ is higher, and the hypothesis $\theta = a$ will be rejected if $P(A \mid a) = 0$ (or is below some relevance threshold). Often, this function is renormalized so that its maximum be equal to 1. We are allowed to let $\pi(a) = P(A \mid a)$ (thanks to this renormalisation) and to interpret this likelihood function in terms of possibility degrees. In particular, it can be checked that $\forall B \subseteq \Theta$, bounds for the value of $P(A \mid B)$ can be computed as:

$$\min_{\theta \in B} P(A \mid \theta) \leq P(A \mid B) \leq \max_{\theta \in B} P(A \mid \theta)$$

which shows that the maxitivity axiom corresponds to an optimistic computation of $P(A \mid B) = \Pi(B)$. It is easy to check that letting $P(A \mid B) = \max_{\theta \in B} P(A \mid \theta)$ is the only way for building a confidence function about Θ from $P(A \mid \theta), \theta \in \Theta$. Indeed, the monotonicity w. r. t. inclusion of the likelihood function $\mathcal{L}$ forces $P(A \mid B) \geq \max_{\theta \in B} P(A \mid \theta)$ to hold [17].

The maximum likelihood principle originally due to Fisher consists in choosing for the value of the parameter, induced by the observation A, $\theta = \theta^*$ that maximizes $P(A \mid \theta)$. It is clear that this selection principle for the estimation of a parameter is in total agreement with possibility theory.

Another element of non-Bayesian statistical analysis is the extraction of a confidence interval for θ on the basis of repeated observations. Let us suppose that the observations $s_1, s_2, \cdots s_k$ result in an estimation $\hat{\theta}$ of the actual value θ^*. Let I_θ be a confidence interval for θ such that $P(I_\theta \mid s_1, s_2, \cdots s_k) \geq 1 - \epsilon$. On can choose the tightest interval E_ϵ of values of θ with probability $1 - \epsilon$, by taking a cut of the density $p(I_\theta \mid s_1, s_2, \cdots s_k)$ (as suggested by the probability-possibility transformations). It is the smallest confidence interval containing the value of θ^* with a confidence level $1 - \epsilon$. One often takes $\epsilon = 0,05$, which is arbitrary. It is clear that letting ϵ varying between 0 and 1, one gets a family of nested sets E_ϵ informing about θ^*. Statistical analysis by means of confidence intervals can thus be understood as the construction of a possibility distribution that provides an imprecise estimate of the value of parameter θ. It can be viewed as an order two possibility distribution over probability measures $P(\cdot|\theta)$.

5 Qualitative uncertainty representations

It seems more natural in an ordinal framework to represent the relative confidence that an agent has between various propositions expressing his/her knowledge rather than trying to force him/her to deliver numerical evaluations. It is indeed easier to assert that a proposition is more credible than another, rather than assessing a belief degree (whose meaning is not always simple to grasp), or even to guess a frequency for each of them. The idea of representing uncertainty by means of relations over a set of events dates back to De Finetti [23], Koopman [80] and Ramsey [98]. They tried to find an ordinal counterpart to subjective probabilities. Later, philosophers of logic such as David Lewis [87] have considered other types of relations, including comparative possibilities in the framework of modal logic. This section offers an overview of ordinal representations of uncertainty, in relation with their numerical counterparts.

The ordinal approaches represent uncertainty by means of a relative confidence relation between propositions (or events) interpreted as subsets of the set S of the states of the world. Such a relation expresses the more or less high confidence of an agent in some propositions rather than in others. Let us denote $\geq_\kappa$ the confidence relation defined on the set of propositions (subsets of S): $A \geq_\kappa B$ means that the agent is at least as confident in the truth of A as in the truth of B. This relation is in general a partial preorder, since the agent may not know the relative confidence between all the propositions. $>_\kappa$ denotes the strict part of $\geq_\kappa$ (i.e., $A >_\kappa B$ if and only if $A \geq_\kappa B$ but not $B \geq_\kappa A$). It expresses that the agent is strictly more confident in A than in B. The agent has equal confidence in A and in B when both $A \geq_\kappa B$ and $B \geq_\kappa A$ hold, which is denoted $A =_\kappa B$. These relations are supposed to satisfy the following properties:

- *Reflexivity* of $\geq_\kappa$: $A \geq_\kappa A, \forall A$;
- *Non-triviality* $S >_\kappa \emptyset$;
- *Coherence with logical deduction.* This is expressed by two properties:
  $A \subseteq B$ entails $B \geq_\kappa A$ (monotony w. r. t. inclusion of $\geq_\kappa$)
  If $A \subseteq B$ and $C \subseteq D$ and $A >_\kappa D$, then $B >_\kappa C$ (well ordered relation);
- *Transitivity* of $>_\kappa$: if $A >_\kappa B$ and $B >_\kappa C$ then $A >_\kappa C$.

These three hypotheses are difficult to challenge. The two coherence conditions w. r. t. deduction are independent except if the relation is complete (i.e. we have $A \geq_\kappa B$ or $B \geq_\kappa A, \forall A, B$) or transitive. Transitivity and completeness of $\geq_\kappa$ become natural if the confidence relation can be represented by a confidence function g with values in $[0, 1]$. In this case, the confidence relation $\geq_\kappa$ is a complete preorder. A confidence function g represents a confidence relation as soon as

$$A \geq_\kappa B \text{ if and only if } g(A) \leq g(B), \forall A, B.$$

All the set functions used for modeling uncertainty (probability measures, possibility measures, belief functions,...) correspond to complete preorders between propositions, satisfying particular properties, if we except the set functions studied by Friedman and Halpern [64] under thename of *plausibility measures*[4] that induce partial preorders of relative confidence.

Comparative probability relations are the first relations of uncertainty that have been introduced (De Finetti [23]; Koopman [80]). They have been studied in detail by Savage [101] in the framework of Decision Theory. A comparative probability relation $\geq_{prob}$ is a complete and transitive confidence relation on the propositions, which satisfies a preadditivity property: if A, B, C are three subsets such as $A \cap (B \cup C) = \emptyset$:

$$B \geq_{prob} C \text{ if and only if } A \cup B \geq_{prob} A \cup C.$$

It is clear that any relation between events induced by a probability measure is preadditive. The converse is false as shown by Kraft et al. [81] by means of the following counter-example on a set S with five elements: Let a comparative probability relation that satisfy the following properties: $s_4 >_{prob} \{s_1, s_3\}$; $\{s_2, s_3\} >_{prob} \{s_1, s_4\}$; $\{s_1, s_5\} >_{prob} \{s_3, s_4\}$; $\{s_1, s_3, s_4\} >_{prob} \{s_2, s_5\}$. The reader can easily check that a comparative probability relation satisfying the above conditions exists, but that there does not exist a probability measure satisfying them. A comparative probability relation is thus an object that is partially non probabilistic, less easy to handle than a probability function. In particular, a probability measure on a finite set is completely defined by the probabilities of the elements, but a comparative probability relation is not fully characterized by its restriction on singletons.

Confidence relations that have this simplicity are possibility and necessity relations. Comparative possibility relations have been independently introduced by David Lewis [87] in the seventies, in the framework of modal logics of counterfactuals, and by Dubois [32] in a Decision Theory perspective. Comparative possibility relations $\geq_\Pi$ are complete and transitive confidence relations satisfying the following characteristic property of *disjunctive stability*:

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

Their numerical counterparts, in the finite setting, are (and only are) the possibility functions Π with values in a totally ordered set L with bottom element 0 and top element 1. Each possibility relation can be entirely specified by means of a unique complete preorder $\geq_\pi$ on the states of the world. $s_1 \geq_\pi s_2$ means that the state s_1 is in general at least as plausible (i.e., normal, not surprizing) as state s_2. The possibility relation on events is then defined as follows :

$$B \geq_\Pi A \text{ if and only if } \exists s_1 \in B, \forall s_2 \in A, s_1 \geq_\pi s_2$$

[4]This name is misleading as they have no relationship to Shafer's plausibility functions.

The degree of possibility of event A thus reflects the plausibility of the state of the world which is the most normal where A is true. The case where the preorder on the events is induced by a partial order on S is studied by Halpern [73]. Possibility relations are not invariant by negation. Comparative necessity relations are defined by duality : $B \geq_N A$ if and only if $\overline{A} \geq_\Pi \overline{B}$. The relation $B \geq_N A$ means that A is at least as certain as B. These necessity relations satisfy a characteristic property called *conjunctive stability*:

$$\forall C, B \geq_N A \text{ entails } C \cap B \geq_N C \cap A$$

The corresponding set functions are necessity measures such that $N(A \cap B) = \min(N(A), N(B))$. Any possibility distribution π from S to a totally ordered set L representing $\geq_\pi$ (i.e. $\pi(s_1) \geq \pi(s_2)$ if and only if $s_1 \geq_\pi s_2$) is defined up to a monotonic transformation. The complete preorder $\geq_\pi$ encodes under a very simple form the generic knowledge of an agent about the relative plausibility of the states of the world. Then one often assumes that for each state $s, \pi(s) > 0$, thus expressing that no state of the world is totally excluded.

Possibility and necessity relations enjoy a remarkable property. Let us call 'belief' any event A such that $A >_\Pi \overline{A}$, then the set of beliefs induced by a possibility relation $\geq_\Pi$ is deductively closed. In particular if A and B are beliefs, the conjunction $A \cap B$ is also a belief. A belief is said to be accepted if an agent accepts to reason as if it is true (and thus to apply the inference rules of classical logic to it). This means that possibility relations account for the notion of accepted belief (Dubois et al. [33]). This property remains when the possibility relation is restricted to a context $C \subseteq S$. One calls 'belief in context C' any event A such that $A \cap C >_\Pi \overline{A} \cap C$. The set of beliefs induced by a possibility relation $\geq_\Pi$ in a context C is also deductively closed. This result relies on the following property of possibility relations (called 'negligibility'): if A, B, C are three disjoint sets,

$$A \cup C >_\Pi B \text{ and } A \cup B >_\Pi C \text{ entails } A >_\Pi B \cup C.$$

This property clearly indicates that $A >_\Pi B$ means that the plausibility of B is negligible w. r. t. the one of A, since in cumulating with B events that are less plausible than A, the plausibility of A is attained. This feature is typical of possibility theory.

There are two main ways for generalizing comparative probability and possibility relations in weakening their characteristic axioms. A first method consists in adopting a restricted form of disjunctive stability, replacing equivalence by implication in the preadditivity axiom: if A, B, C are three subsets such as $A \cap (B \cup C) = \emptyset$:

$$B \geq_\kappa C \text{ entails } A \cup B \geq_\kappa A \cup C. \tag{34}$$

The results proved in Dubois [32] and Chateauneuf [12] show that the class of set functions captured by the weak preadditivity axiom (34) exactly contains the pseudo-additive (or decomposable) confidence functions g, i.e.,which are such that there exists an operation $\oplus$ on the codomain of g such that for each pair of disjoint subsets $A, B, g(A \cup B) = g(A) \oplus g(B)$. The cases where $\oplus = \max$ and $\oplus = +$ cover possibility and probability measures, respectively.

The other extension consists in restricting the scope of the weak preadditivity axiom to subsets A, B, C such as $A \cap (B \cup C) = \emptyset$ and $C \subseteq B$. Any relative confidence relation $\geq_\kappa$ obeying this restriction of the preadditivity axiom is representable by a plausibility function in the sense of Shafer (see Wong et al. [124]).

6 Conditioning in non-additive representations

The generalisation of the notion of probabilistic conditioning to other theories of uncertainty is not straightforward for at least two reasons:

- As pointed out in section 3.2, probabilistic conditioning is often directly defined as a ratio of two quantities and not as the probability of a genuine conditional event. However, splitting the conditional event from the probability measure, one may better understand how to generalise the notion of conditioning.
- Probabilistic conditioning has been used for several types of very different tasks: learning from observations, prediction from observations, and the revision of uncertain information. Moreover, there are several ways of formally generalising the probabilistic conditioning. It is not obvious that the various tasks can be modelled by the same form of conditioning.

First a clarification is in order. The quantity $P(A \mid C)$ is often presented as the probability of event A when C is true. The example below shows that it is a misconception.

> **Example**: Let us consider balls drawn from a bag S containing five balls numbered from 1 to 5. It is clear that $P(even \mid \{1,2,3\}) = P(even \mid \{3,4,5\}) = \frac{1}{3}$. If one understands these results as: *if the ball is in* $\{1,2,3\}$, *then the probability that it is even is* $\frac{1}{3}$ and *if the ball is in* $\{3,4,5\}$, *then the probability that it is even is* $\frac{1}{3}$, one is logically led to conclude that the probability that the ball is even is $\frac{1}{3}$ *in any case* since $S = \{1,2,3\} \cup \{3,4,5\}$. However $P(Pair \mid S) = \frac{2}{5}$.

The reason for this paradox, is a misinterpretation of the conditional probability $P(A \mid C)$. In fact, this is the probability of A when one does not know anything else than the truth of C (in the example: if one only knows that the number of the ball is in the set $\{1,2,3\}$). Note that "*knowing only that the ball is in* $\{1,2,3\}$ *or that the ball is in* $\{3,4,5\}$" is not equivalent to knowing nothing. Thus, one should understand $P(A \mid C)$ as the probability of an event $A \mid C$ which involves a non classical implication, different from the material one $\overline{C} \cup A$, since in general it is false that $P(A \mid C) = P(\overline{C} \cup A)$, and moreover it is not true that $P(A \mid C) \leq P(A \mid C \cap B)$ (lack of monotonicity) while of course $P(\overline{C} \cup A) \leq P(\overline{C \cap B} \cup A)$.

Besides, it is important to distinguish the prediction problem from the revision problem. When dealing with prediction, we have at our disposal a model of the world under the form of probability distribution P issued for instance from a representative set of statistical data. This is what we call 'generic information' or 'generic knowledge' (for instance, medical knowledge synthesized by causal relations between diseases and symptoms). Assume we have some observations on the current state of the world, i.e. a particular situation, what we call singular information, under the form of a proposition C (e.g. some medical test results for a patient). Then, one tries to formulate some statements A about the current world with their associated degrees of belief (e.g. predict the disease of the patient). Then the conditional probability $P(A \mid C)$ (which is for instance the frequency of observation of A in context C) is used for estimating a degree of belief that the current world satisfies A.

The revision scenario is different: given a probability distribution P (which may represent generic information or not), one learns that the probability of an event C is 1 (and not $P(C) < 1$ as it was supposed before). Then the problem is to determine the new probability measure P', such that $P(C) = 1$, which is the closest to P in some sense, in order to comply with a minimal change principle. Then, it can be shown that if we use an appropriate relative information measure, it follows that $P'(A) = P(A \mid C), \forall A$ [123].

Note that, in the prediction problem, generic knowledge remains unaffected by singular evidence, which is handled apart. Finally, learning can be viewed as bridging the gap between generic and singular information. Bayes theorem is instrumental to let prior knowledge be altered by singular evidence, when checking the validity of predictions. An important problem is to see what remains of Bayesian learning when prior knowledge is incomplete. While the answer to this question is not well-understood yet, the imprecise Dirichlet model [4] provides some insight on this problem for imprecise probabilities. For belief functions, little has be done as it is a theory of handling singular uncertain evidence, and not so much an extension of Bayesian probabilistic modeling. In the following, we focus on prediction, revision and later on the fusion of evidence.

6.1 Conditional events and qualitative conditioning

De Finetti [22] was the first to regard the conditional probability $P(A \mid C)$ as the probability of a *tri-event* $A \mid C$ that should be read "if what is known is described by C then conclude A", where A and C represents classical propositions (interpreted as subsets of S). A tri-event $A \mid C$ partitions the set of states $s \in S$ into three subsets:

- either $s \in A \cap C$; then s is said to be an *example* of the rule "if C then A". The tri-event is then true (value 1) at s;
- or $s \in \overline{A} \cap C$; then s is said to be a *counter-example* of the rule "if C then A". The tri-event is then false (value 0) at s;
- or $s \in \overline{C}$; then s is said to be *irrelevant* for the rule "if C then A", i.e. the rule does not apply to s. The tri-event then takes a third truth value (I) at s.

The third truth value can be interpreted in various ways. According to Goodman et al. [71], it corresponds to an hesitation between true and false, i.e. $I = \{0, 1\}$. This is philosophically debatable but suggests the equivalence between a tri-event and a family of subsets of S, lower bounded by $A \cap C$ (this is the case when we choose $I = 0$) and upper bounded by $\overline{C} \cup A$ representing material implication (this is the case when we choose $I = 1$). It is easy to check that any subset B such as $A \cap C \subseteq B \subseteq \overline{C} \cup A$ satisfies the identity $A \cap C = B \cap C$. Thus, one has an Bayesian-like equality of the form:

$$A \cap C = (A \mid C) \cap C. \tag{35}$$

as this identity is valid for any representative of the family $\{B : A \cap C \subseteq B \subseteq \overline{C} \cup A\}$. This family is an interval in the algebra of subsets of S, fully characterized by the nested pair $(A \cap C, \overline{C} \cup A)$.

The third truth-value I may be also seen as really expressing "inapplicable" [8], which underlies the definition of a conjunction of conditional events by means of a truth table with three values, in a non-monotonic three-valued extension of propositional logic (Dubois and Prade [49]). Lastly, for De Finetti and his followers [16], the truth value I should be changed into the probability $P(A \mid C)$. Indeed,the probability $P(A \mid C)$ is then seen as the price of a lottery ticket in a conditional bet that yields, if condition C is satisfied, 1 euro when A takes place, 0 when A does not take place, and where the price paid is reimbursed (the bet is called off) if condition C (which is the precondition for the game to take place) turns to be false.

Relation (35) is the Boolean variant of Bayes equation $P(A \cap C) = P(A \mid C) \cdot P(C)$. Moreover $P(A \mid C)$ is indeed a function of $P(A \cap C)$ and $P(\overline{C} \cup A)$ only, since (if $P(C) > 0$):

$$P(A \mid C) = \frac{P(A \cap C)}{P(A \cap C) + 1 - P(\overline{C} \cup A)}. \tag{36}$$

Thus, it is possible to separate the tri-event from the conditional probability. Therefore they are two ways of generalizing the probabilistic conditioning to confidence functions g that differ from probabilities.

- Either one states that $g(A \cap C)$ only depends on $g(A \mid C)$ and $g(C)$, via a function ϕ. This is the approach followed by Cox (see Paris [96]). The constraints induced by the Boolean algebra of events, together with some natural technical conditions such as the strict increasingness of ϕ enforce $g(A \cap C) = g(A \mid C) \cdot g(C)$ in practice.
- Or, the conditional measure $g(A \mid C)$ is directly defined by replacing P by g in (36).

The equivalence between the two approaches, which holds for probabilities, is no longer true for more general set functions. In the case of non-numerical possibility theory, with possibility values on a finite scale L, only the first option, generalizing (35) is possible. Then, we state

$$\Pi(A \cap C) = \min(\Pi(A \mid C), \Pi(C)). \tag{37}$$

This equation has no unique solution. Nevertheless, in the spirit of possibility theory, one is led to select the least informative solution, i.e., for $C \neq \emptyset$, and $A \neq \emptyset$:

$$\Pi(A \mid C) = 1 \text{ if } \Pi(A \cap C) = \Pi(C), \text{ and } \Pi(A \cap C) \text{ otherwise.} \tag{38}$$

This is similar to conditional probability, but there is no longer any division of $\Pi(A \cap C)$. If $\Pi(C) = 0$, then $\Pi(A \mid C) = 1$ provided that $A \neq \emptyset$. Conditioning by an impossible event destroys information.

The conditional necessity measure is then defined by $N(A \mid C) = 1 - \Pi(\overline{A} \mid C)$. It coincides with the necessity of the material implication except if $\Pi(A \cap C) = \Pi(C)$. Note that the dual equation $N(A \cap C) = \min(N(A \mid C), N(C))$ is not very interesting, since its minimal solution is $N(A \mid C) = N(A \cap C) = \min(N(A), N(C))$, which comes down to stating $\Pi(A \mid C) = \Pi(\overline{C} \cup A)$. On the other hand, the solution of equation (37) captures ordinal conditioning of the previous section, since it can be checked that $N(A \mid C) > 0 \iff \Pi(A \cap C) > \Pi(\overline{A} \cap C)$ when $\Pi(C) > 0$. This means that a proposition A is accepted as true in context C if it is more plausible than its negation in this context. The non monotonic nature of this type of conditional possibility can be seen by noticing that we may have both $N(A \mid C) > 0$ and $N(\overline{A} \mid B \cap C) > 0$, i.e., the arrival of information B may lead to reject proposition A, which was accepted before in context C. See Benferhat et al. [3] for a more detailed study of non-monotonicity in this framework.

6.2 Conditioning for belief functions and imprecise probabilities

Most of the time, the information encoded by a probability distribution refers to a population (the set of situations that correspond to the results of the statistical tests). This is a form of generic information, typically frequentist. This information can be used for inferring beliefs about a particular situation for which we have incomplete but clear-cut observations. This is called prediction. If $P(A \mid C)$ is the (frequentist) probability of having A in context C, the confidence of the agent in proposition A, when he/she knows information C, is estimated by quantity $P(A \mid C)$, assuming that the current situation is typical of environment C. The belief of the agent in proposition A in the current situation changes from $P(A)$ to $P(A \mid C)$ when it has been observed that C is true in the current situation and nothing else. Conditioning here is used for updating the beliefs of the agent about the current situation by exploiting generic information. In the example used above, the probability measure P represents the medical knowledge (often compiled under the form of a Bayesian network). The singular information C represents the results of tests for a patient. $P(A \mid C)$ is the probability of having disease A for patients for whom C has been observed; this value also estimates the singular probability (belief) that this patient has this disease. Note that in this type of inference, the probability measure P does not change, only singular beliefs change. One only applies the available generic knowledge to a reference class C, what is called *focusing* in (Dubois, Moral, and Prade [37]).

When probability P is subjective, it may have a singular nature as well (when betting on the occurrence of a non-repetable event). In this case conditioning can be interpreted as an updating of a singular probability *by a piece of information of the same nature*. In this case, information C is interpreted as $P(C) = 1$, which represents a constraint that has to be taken into account when revising P. For instance [54], in a criminal affair where the investigator suspects Peter, Paul and Mary with probabilistic confidence degrees $\frac{1}{4}$, $\frac{1}{4}$ and $\frac{1}{2}$ respectively, and then learns that Peter has an alibi, i.e. $P(\{\text{Mary, Paul}\}) = 1$. We then have to revise these singular probabilities. The use of conditional probability for handling this revision of the probabilities is often proposed (and justified by the minimal change principle, already mentioned above), which yields probabilities $\frac{1}{3}$ and $\frac{2}{3}$ for Paul and Mary respectively. However, the problem of revising P is different from the one of updating singular beliefs on the basis of generic information.

Lastly, one may also want to justify the revision of a frequentist probability after the occurrence of major events. In the example of the opinion poll about a future election, let suppose that for each candidate, his/her frequentist probability to be elected has been obtained (i.e. everybody supplied a precise favourite candidate). Suppose now that a candidate withdraws. What becomes of the probabilities? Applying Bayesian conditioning in this situation is questionable, since it assumes that the votes of the electors that previously supporting the withdrawn candidate are transferred to the other candidates proportionally to the number of the potential votes previously estimated. It would be more convincing to make the assumption that the transfers will be done towards the nearest neighbors of the withdrawn candidate in terms political affinity (which corresponds to the 'imaging' rule proposed by Lewis ([86]). This case questions the alleged universality of Bayesian conditioning, even for probabilities. In such a situation, it would be even better to run the opinion poll again.

In the case where the generic knowledge of the agent is represented by imprecise probabilities, Bayesian plausible inference is generalized by performing a sensitivity analysis on the conditional probability. Let $\mathcal{P}$ be a family of probability measures on S. For each proposition A a lower bound $P_*(A)$ and an upper bound $P^*(A)$ of the probability degree of A are known. In presence of singular observations summarized under the form of a context C, the belief of an agent in a proposition A is represented by the interval $[P_*(A \mid C), P^*(A \mid C)]$ defined by

$$P_*(A \mid C) = \inf\{P(A \mid C), P(C) > 0, P \in \mathcal{P}\}$$

$$P^*(A \mid C) = \sup\{P(A \mid C), P(C) > 0, P \in \mathcal{P}\}.$$

It may happen that the interval $[P_*(A \mid C), P^*(A \mid C)]$ is larger than $[P_*(A), P^*(A)]$, which corresponds to a loss of information in specific contexts. This property reflects the idea that the more singular information is available about a situation, the less informative is the application of generic information to it (since the number of statistical data that fit this situation may become very small). We see that this form of conditioning does not correspond at all to the idea of enriching generic information, it is only a matter of querying it.

Belief and plausibility functions in the sense of Shafer [104] are mathematically speaking important particular cases of lower and upper probabilities, although these functions were independently introduced, without any reference to the idea of imprecise probability. Information is supposed to be represented by the assignment of non-negative weights $m(E)$ to subsets E of S. In a generic knowledge representation perspective, $m(E)$ is, for instance, the proportion of imprecise results, of the form $x \in E$, in a statistical test on a random variable x. In this framework, plausible inference in context C consists in evaluating the weight function $m(\cdot \mid C)$ induced by the mass function m on the set of states C, taken as the new frame. Three cases should be considered:

- $E \subseteq C$: In this case, $m(E)$ remains assigned to E.
- $E \cap C = \emptyset$: In this case, $m(E)$ no longer matters and is eliminated.
- $E \cap C \neq \emptyset$ and $\overline{E} \cap C \neq \emptyset$: In this case, some fraction $\alpha_E \cdot m(E)$ of $m(E)$ remains assigned to $E \cap C$ and the rest, i.e. $(1 - \alpha_E) \cdot m(E)$, is allocated to $\overline{E} \cap C$. But this sharing process is unknown.

The third case corresponds to incomplete observations E that neither confirm nor disconfirm C. We have not enough information in order to know if, in each of the situations corresponding to these observations, C is true or not, since only E is known. Suppose that the values $\{\alpha_E, E \subseteq S\}$ were re known. It is always known that $\alpha_E = 1$ and $\alpha_E = 0$ in the first and second cases respectively. Then, we can build a mass function $m_\alpha^C(\cdot)$. Note that a renormalisation of this mass function is necessary, in general, as soon as $Pl(C) < 1$ (letting $m_\alpha(\cdot \mid C) = \frac{m_\alpha^C(\cdot)}{Pl(C)}$). If one denotes by $Bel_\alpha(A \mid C)$ and $Pl_\alpha(A \mid C)$ the belief and plausibility functions obtained by focusing on C, based on the allocation vector α, the conditional belief and plausibility degrees on C are defined by

$$Bel(A \mid C) = \inf_\alpha Bel_\alpha(A \mid C),$$

and

$$Pl(A \mid C) = \sup_\alpha Pl_\alpha(A \mid C)$$

One still obtains belief and plausibility functions (Jaffray [75]), and necessity and possibility measures if we start with such measures (Dubois and Prade [48]). The following results show that what is obtained is a generalization of Bayesian inference:

$$Bel(A \mid C) = \inf\{P(A \mid C) : P(C) > 0, P \geq Bel\} = \frac{Bel(A \cap C)}{Bel(A \cap C) + Pl(\overline{A} \cap C)}$$

$$Pl(A \mid C) = \sup\{P(A \mid C) : P(C) > 0, P \geq Bel\} = \frac{Pl(A \cap C)}{Pl(A \cap C) + Bel(\overline{A} \cap C)}$$

It is easy to see that $Pl(A \mid C) = 1 - Bel(\overline{A} \mid C)$, and that these formulas generalize probabilistic conditioning under form (36): $Bel(A \mid C)$ is indeed a function of $Bel(A \cap C)$ and of $Bel(\overline{C} \cup A)$ (and similarly for $Pl(A \mid C)$). Note that if $Bel(C) = 0$ and $Pl(C) = 1$ (complete ignorance regarding C) then all the focal sets of m overlap C without being contained in C. In this case, $Bel(A \mid C) = 0$ and $Pl(A \mid C) = 1, \forall A \neq S, \emptyset$: one cannot infer anything in context C.

The other conditioning, called 'Dempster conditioning', proposed by Shafer [104] and Smets [116], systematically assumes $\alpha_E = 1$ as soon as $E \cap C \neq \emptyset$. It supposes a transfer of the full mass of each focal set E to $E \cap C \neq \emptyset$ (followed by a renormalisation). This means that we interpret the new information C as modifying the initial belief function in such a way that $Pl(\overline{C}) = 0$: situations where C is false are considered as impossible. If one denotes $Pl(A \mid\mid C)$ the plausibility function after revision, we have:

$$Pl(A \mid\mid C) = \frac{Pl(A \cap C)}{Pl(C)}$$

This constitutes another generalisation of probabilistic conditioning in the sense of equation (35). The conditional belief is then obtained by duality $Bel(A \mid\mid C) = 1 - Pl(\overline{A} \mid\mid C)$. Note that with this conditioning, the size of focal sets diminishes, thus information becomes more precise, and the intervals $[Bel, Pl]$ become more tighter (they are always tighter than those obtained by focusing). Dempster conditioning thus corresponds to a process where information is enriched, which contrasts with focusing. If $Bel(C) = 0$ and $Pl(C) = 1$ (complete ignorance about C), conditioning on C in the sense of Dempster rule significantly increases the precision of resulting beliefs.

In the more general framework of imprecise probabilities, a blind application of revision by a piece of information C consists in adding the supplementary constraint $P(C) = 1$ to the family $\mathcal{P}$, i.e.

$$P_*(A \mid\mid C) = \inf\{P(A \mid C), P(C) = 1, P \in \mathcal{P}\};$$

$$P^*(A \mid\mid C) = \sup\{P(A \mid C), P(C) = 1, P \in \mathcal{P}\}.$$

But, it may happen that the set $\{P \in \mathcal{P},\ P(C) = 1\}$ is empty (it is always the case in the classical Bayesian framework since $\mathcal{P}$ is a singleton). One then applies the maximal likelihood principle (Gilboa and Schmeidler,[67]) and we replace the condition $P(C) = 1$ by $P(C) = P^*(C)$ in the above equation. In this way, we generalize Dempster rule (which is recovered if P^* is a plausibility function).

This type of conditioning has nothing to do with the previously described focusing problem, since in the view of Shafer and Smets, the mass function m does not represent generic knowledge, but rather uncertain singular information (non fully reliable testimonies, more or less certain clues) collected about a particular situation. These authors consider a form of reasoning under uncertainty where generic knowledge is not taken into account, but where all the pieces of information are singular. In the crime example, suppose that the organizer of the crime tossed a coin for deciding whether a man or a woman is recruited to be the killer. This piece of uncertain singular information is represented by the mass function $m(\{\text{Peter, Paul}\}) = \frac{1}{2}$ (there is no information available about Peter alone and Paul alone), and $m(\{\text{Mary}\}) = \frac{1}{2}$. Now, if we learn that Peter has an alibi, the focal set {Peter, Paul} reduces to {Paul} and we deduce after revision, that $P(\{\text{Mary}\}) = P(\{\text{Paul}\}) = \frac{1}{2}$. Note that the Bayesian approach would split the mass $m(\{\text{Peter, Paul}\})$ equally between Peter and Paul. Bayesian conditioning then yields $P(\{\text{Mary}\}) = 2 \cdot P(\{\text{Paul}\}) = \frac{2}{3}$, which may sound debatable when dealing with uncertain singular pieces of information (let alone at a court of law).

7 Fusion of imprecise and uncertain information

The problem of fusing distinct pieces of information coming from different sources has become increasingly important in several areas, such as robotics (multi-sensor fusion), image processing (merging of several images), risk analysis (expert opinions fusion), or databases (fusion of knowledge bases). However, fusion has received little attention in the probabilistic tradition. In the frequentist view, one works with a unique probability distribution issued from a set of observations. In the subjectivist tradition, one often considers that uncertainty is expressed by a unique agent. In the last thirty years, the problem of fusing pieces of information has emerged as a fundamental issue when representing information coming from several sources. The fusion of pieces of information differs from the fusion of multiple criteria or multiple agent preferences. In that latter case one usually looks for a compromise between points of views or agents. Each agent may be led to accept options that he/she had not proposed at the beginning. In contrast, the aim of information fusion is to lay bare what is true among a collection of data that are often imprecise and inconsistent. Consequently, the operations that are natural for fusing different pieces of information are not necessarily those needed for fusing preferences. The fusion problem can be stated in similar terms independently from the representation of uncertainty that is used: in each uncertainty theory, one can find the same fusion modes, even if they are expressed by different operations. Besides, the fusion problem differs from the revision of information upon the arrival of a new piece of information (which is based on the notion of conditioning). The fusion problem is by nature symmetrical: sources play similar role even if they may be (are often) heterogeneous. This contrasts with revision, where prior information is minimally changes on the basis of new information. When fusing pieces of information, there may be no prior knowledge available, and if any, it is modified by the pieces of information coming from several sources in parallel.

In its simplest form a fusion problem can be stated as follows, when the sources provide incomplete pieces of information: Assume there are two sources 1 and 2 that inform us about the value of a variable x taking its value in S. According to the first source, $x \in A_1$, while according to the second one, $x \in A_2$. The fusion problem consists in deducing the most useful plausible information contained in what sources delivered. It is obvious that the result should depend on hypotheses on the quality of the sources. There are three kinds of assumptions:

1. *The two sources are reliable.* One concludes that $x \in A_1 \cap A_2$. This reasoning presupposes that the pieces of information that we start with are coherent. If $A_1 \cap A_2 = \emptyset$, then the hypothesis that the two sources are reliable no longer holds.

2. *At least one of the two sources is reliable.* One concludes that $x \in A_1 \cup A_2$. This reasoning no longer presupposes that the pieces of information that we start with are coherent. Thus if $A_1 \cap A_2 = \emptyset$, one can still deduce a non trivial piece of information (except if x is an all-or-nothing variable). But there is an important loss in precision.

3. *The two sources are identical and provide independent information.* In this case one can consider that $A_1 \cap A_2$ is the set of values that are the most plausible (since both sources declare them as feasible), while the values in $(A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2)$ are less plausible, but not excluded (since, at least one of the two sources declare them as possible).

These three kinds of combination can be found in all formalisms. The first one is called *conjunctive fusion*, since it performs the intersection of the sets of values that are possible for x according to each source. This is the usual fusion mode in classical logic. If several propositions of the form $x \in A_i$ are asserted as true, the values resulting from the combination are the ones for which all the propositions are true. The second is called *disjunctive fusion.* It corresponds to a classical mode for dealing with inconsistency in logic (Rescher and Manor [100]): If the propositions of the form $x \in A_i$ are contradictory, then one looks for maximal consistent subsets of propositions, assuming that reality corresponds to one of these subsets (here reduced to $\{A_1\}$ and $\{A_2\}$). The third mode is of another nature: the hypothesis of independence of the sources allows for a counting process. For each value of x, one counts the number of sources that do not exclude it, this number reflects the plausibility of each of these values. This is typically what is done in statistics, but in that latter case each observation is supposed to be precise ($x = a_i$) and comes from the same unique aleatory source in an independent way. Moreover, it is also supposed that many more than two observations are reported. Up to these remarks, collecting statistical data agrees with the third fusion mode, which may be thus termed *cumulative fusion.* In the above elementary case, it can be expressed by the arithmetic mean of the characteristic functions of A_1 and A_2, or yet under the form of a mass distribution such that $m(\{A_1\}) = m(\{A_2\}) = \frac{1}{2}$.

In the following, we explain how these three modes of fusion can be expressed in the different uncertainty formalisms studied in this paper: probabilities, possibilities, belief functions, imprecise probabilities. We shall see that the closure condition, which supposes that when one fuses pieces of information that are expressed in a given formalism, the combination result should be also expressible in this formalism, may be problematic. Requiring this condition may forbid some fusion modes. For instance, it is clear that cumulative fusion does not preserve the all-or-nothing nature of the pieces of information in the above elementary case, contrary to the situations for conjunctive or disjunctive fusions. In order to define all fusion modes in all formalisms, we shall see that the result of the fusion brings us from a particular setting to a more general one (for example from possibilities or probabilities to belief functions).

7.1 Non-Bayesian probabilistic fusion

It is supposed that source i provides a probability measure P_i on S. One looks for a function f, which is non-decreasing, monotonic, from $[0,1]^n$ to $[0,1]$ such that the set function $P = f(P_1, \ldots, P_n)$ is still a probability measure. Existing results show that, under mild conditions, especially, $f(0, \ldots, 0) = 0$ and $f(1, \ldots, 1) = 1$, the only possible fusion function is the weighted average [84], i.e.

$$\forall A \subseteq S, f(P_1(A), \ldots, P_n(A)) = \sum_{i=1}^{n} \alpha_i \cdot P_i(A),$$

where $\sum_{i=1}^{n} \alpha_i = 1$, with $\alpha_i \geq 0, \forall i$. This amounts to requiring that aggregation commutes with marginalization.

This is a cumulative fusion mode. It considers the sources as independent aleatory generators of precise values, the weight α_i reflects the number of observations produced by source i. In the framework of expert opinion fusion, it is supposed that each expert produces a probability measure expressing what he/she knows about the value of a parameter x, the weight α_i reflecting the reliability of expert i, understood as the probability that expert i is "the right one". These weights are estimated by testing the expert on questions, the answer to which is supposedly known (Cooke [18]).

This is the only fusion mode allowed by this approach. One may also fuse probability densities by means of other operations such as the geometric mean, provided that the result is renormalized, which broadens the spectrum of possible fusion operations (French [63]). But the commutation with marginalization operation is then lost.

7.2 Bayesian probabilistic fusion

Another approach to the fusion problem presupposes that sources provide precise evaluations $x_1, \ldots, x_n$ for the value of x, but that these evaluations are inexact. The probability $P(x_1, \ldots, x_n \mid x = s_j)$ that sources jointly provide the n-tuple of values $x_1, \ldots, x_n$ when the real value of x is equal to s_j is supposed to be known. This information models the joint behavior of the sources. Moreover, prior probabilities p_j that $x = s_j, \forall j$ are also supposed to be available. Under these hypotheses, one can compute, by means of Bayes theorem, the probability $P(x = s_j \mid x_1, \ldots, x_n)$ that the real value of x is equal to s_j when each source i provides a value x_i:

$$P(x = s_j \mid x_1, \ldots, x_n) = \frac{P(x_1, \ldots, x_n \mid s_j) \cdot p_j}{\sum_{k=1}^{n} P(x_1, \ldots, x_n \mid s_k) \cdot p_k}. \quad (39)$$

In spite of its appeal, this approach is very demanding in pieces of information. The probability $P(x_1, \ldots, x_n \mid x = s_j)$ reflects the dependency between sources. It is seldom available, since it requires a great number of values to be specified. In practice, it is easier to obtain the marginal probabilities $P(x_i \mid x = s_j)$ that each source i provides the value x_i when the real value of x is equal to s_j. By default, sources are assumed to be conditionally independent of the true value of x, which gives:

$$P(x = s_j \mid x_1, \ldots, x_n) = \frac{P(x_1 \mid s_j) \cdot \cdots \cdot P(x_n \mid s_j) \cdot p_j}{\sum_{k=1}^{n} P(x_1, \ldots, x_n \mid s_k) \cdot p_k}. \quad (40)$$

Besides, we need a prior probability about the value sources are supposed to provide information on. Such prior information is often missing, since if it were available, one might not even need the pieces of information provided by sources. In practice, one is obliged to provide a subjective estimate of the prior probability (taken as uniform by default), which may influence the result. Nevertheless, let us remark that this fusion mode is conjunctive (since the product of the likelihood functions is performed). One might think of defining a disjunctive fusion mode by computing $P(x = s_j \mid x_1$ or $\ldots$ or $x_n)$, the probability that the real value for x is equal to s given at least one of the values provided by the sources.

7.3 Fusion in possibility theory

In this framework, each source i is supposed to provide a possibility distribution π_i defined on S. For fusing such pieces of information, the whole panoply of fuzzy set aggregation operations is available (Dubois and Prade [43], Chap. 2; Fodor and Yager [62], Calvo and Mesiar [9]). In particular, the three basic information fusion modes can be expressed and a resulting possibility distribution π on S by obtained under the form $\pi(s) = f(\pi_1(s), \ldots, \pi_n(s)), \forall s \in S$ for an appropriate operation f.

For conjunctive fusion, one can use triangular norms (Klement et al. [78]) which are semi-group operations of the unit interval (hence associative), monotonically increasing, and with neutral element 1. The main operations of this kind are the minimum operation, the product, and the linear conjunction $\max(0, a + b - 1)$. The advantage of choosing

$$\pi(s) = \min(\pi_1(s), \ldots, \pi_n(s)), \forall s \in S$$

is the idempotency of this operation. If all sources provide the same distribution π, it is this distribution that is taken as result. This property enables us to cope with the case where the sources are redundant (for example when experts have the same background knowledge), without requiring any assumption about the independence of the sources. But, if one is sure that the information sources are independent, it may be desirable to have a reinforcement effect (what all the sources consider to have a low plausibility will receive a very low global plausibility). This effect is captured by the product: $\pi(s) = \pi_1(s) \cdot \pi_2(s) ... \cdot \pi_n(s), \forall s \in S$.

The reinforcement effect obtained with the linear conjunction $\max(0, a + b - 1)$ is much stronger, since values that are considered as being very little plausible, but not impossible, by all sources are eliminated after the fusion. In fact, this operation applies when it is known that a certain number k of sources lie (Mundici [94]). Then an information item of the form $x \in A$ proposed by a source is modelled as $\pi_i(s) = 1$ if $s \in A$ and $1 - \frac{1}{k+1}$ otherwise, which is all the higher as k is large. The linear conjunction enables to eliminate for sure values at least $k + 1$ sources declare impossible.

All these operations clearly generalize the conjunctive fusion of two all-or-nothing pieces of information and presuppose that the possibility distributions provided by the sources are not contradictory. Nevertheless, the resulting possibility distribution will be often sub-normalized ($\pi(s) < 1, \forall s \in S$). The quantity $Cons = \max_{s \in S} \pi(s)$ measures the degree of consistency between the sources. The fusion result may be renormalized if it is sure that the sources are reliable (since the true value of x is among the values that are not eliminated by any source, even if its possibility is very low). When sources are independent one gets

$$\pi(s) = \frac{\pi_1(s) \cdot \cdots \cdot \pi_n(s)}{\max_{s \in S} \pi_1(s) \cdot \cdots \cdot \pi_n(s)}, \forall s \in S. \tag{41}$$

Renormalisation preserves associativity if the combination operation is the product. But, when renormalisation is applied to minimum operation, associativity is lost (Dubois and Prade [44]). Let us notice the striking similarity between Bayesian fusion (40) and possibilistic fusion(41), especially when letting $\pi_i(s) = P(x_i \mid x = s)$, which has been justified above. The difference between the two fusion operations lies in the presence of the prior probability in (40), and in the type of renormalisation (probabilistic or possibilistic). The two resulting distributions are even proportional if a uniform prior probability is chosen in (40). This coincidence between Bayesian and possibilistic approaches indicates their mutual coherence, and confirms the conjunctive nature of Bayesian fusion. However, the similarity of numerical results should not hide a serious difference at the interpretation level. In the probabilistic framework, it is supposed that the posterior probability of each possible value of x can be computed in a precise way. In the possibilistic framework, and in agreement with the non Bayesian probabilistic tradition, fusion only provides a likelihood degree for the possible values of x. This information is poorer than a probability degree, which is too rich information in the case of partial ignorance.

When the value of the consistency index is too low, renormalization makes the conjunctive fusion numerically unstable [44]. Inconsistency is all the more likely as the number of sources is high. In this case disjunctive fusion becomes more appropriate and relevant. For that latter fusion mode, triangular co-norms (Klement et al. [78]) can be used. They are semi-groups of the unit interval, monotonically increasing, and with neutral element 0. Co-norms u are obtained by De Morgan duality from triangular norms t under the form $u(a, b) = 1 - t(1-a, 1-b)$. The main operations of this kind for disjunctive fusion are the maximum operation, the probabilistic sum $a + b - ab$, and the bounded sum $\min(1, a+b)$. This type of fusion operation does not require any renormalisation step. But, since it supposes only that one source is reliable, the obtained result may be very imprecise, in particular if the number of sources is high, due to the higher risk of scattered pieces of information. It is then possible to use milder fusion modes. For instance, a quantified fusion may be used: It is assumed that there are k reliable sources among n, then first a conjunctive fusion is performed inside each group of k sources, and these partial results are then combined disjunctively. One may optimize the value of k by trying to maximize the informativeness of the result (in order to choose k not too small), while minimizing inconsistencies (choosing k not too large) (see Dubois and Prade [50]). One may also look for maximal sub-groups of sources that are together consistent, then perform conjunctive fusion inside these groups, and finally combine these partial results disjunctively (Dubois and Prade [52]). This can be done for cuts of each possibility distribution, which no longer leads to a possibility distribution for the result, but a belief function [28].

Lastly, one may also apply a cumulative fusion mode to possibilistic pieces of information, under the form of a weighted arithmetic mean, $\sum_{i=1}^{n} \alpha_i \cdot \pi_i$, when the sources are numerous and independent. Nevertheless, the convex combination of possibility measures is not a possibility measure, but again a belief function, since the consonance of focal sets is not preserved by convex sum. Only the disjunctive fusion of possibility measures based on the maximum operation provides a possibility measure [45].

7.4 Fusion of belief functions

It is now supposed that the two sources provide two mass functions m_1 and m_2 defined on frame S. Shafer [104] has proposed a conjunctive combination rule that may be related to the Bayesian fusion method and generalizes set intersection. It amounts to performing the intersection of each focal set A_1 of m_1 with each focal subset A_2 of m_2 and to allocate mass $m_1(A_1) \cdot m_2(A_2)$ to the subset $A_1 \cap A_2$ (which may be empty). In order to obtain a normal mass function, the result is renormalized by dividing by the sum of masses allocated to non-empty subsets. Thus, it leads to an associative combination rule:

$$\forall A \subseteq S, m(A) = \frac{\sum_{A_1,A_2:A=A_1\cap A_2} m_1(A_1) \cdot m_2(A_2)}{\sum_{A_1,A_2:A_1\cap A_2\neq\emptyset} m_1(A_1) \cdot m_2(A_2)} \tag{42}$$

It is easy to check that this fusion rule is also commutative, but non-idempotent. This rule supposes that the information *sources* (not the underlying variables) are independent. The normalisation factor is an evaluation of the consistency between the sources. One may also notice that the plausibility function Pl induced by m, restricted to the singletons in S satisfies the following property: $\forall s, Pl(\{s\})$ is proportional to the product $Pl_1((\{s\})) \cdot Pl_2((\{s\}))$ (equality holds if the sources are consistent, i.e., $\forall A_1, A_2 : A_1 \cap A_2 \neq \emptyset$).

Appliying this combination rule to two possibility measures Π_1 and Π_2, it can be seen that the resulting mass function is not necessarily consonant (if the focal sets A_1 of m_1 and A_2 of m_2 are nested, it may be not the case for the subsets of the form $A_1 \cap A_2$). Nevertheless, the possibilistic fusion rule (41) is an approximation of Dempster rule in this case, since it provides a possibility distribution that is proportional to $Pl(\{s\})$. This fusion rule may be also applied to probability distributions p_1 and p_2. It amounts to performing products $p_1(s) \cdot p_2(s), \forall s \in S$ and renormalizing the distribution thus obtained. If one combines a mass function m_1 with a probability function p, what is obtained is a probability distribution proportional to $p(s) \cdot Pl_1(s)$. Combining three mass functions m_1, m_2 and m_3 by Dempster rule, where the last one is a probability ($m_3 = p$), it amounts to apply the Bayesian fusion rule to sources 1 and 2, viewing $Pl_1(\{s\})$ and $Pl_2(\{s\})$ as likelihood functions and m_3 as a prior probability.

Dempster rule is also numerically unstable when the sources are not very consistent, i.e., when the renormalisation factor in (42) is small [44]. In this case, one may use the disjunctive counterpart to Dempster rule, which amounts to replacing intersection by union in formula (42), i.e., [41] :

$$m(A) = \sum_{A_1,A_2:A=A_1\cup A_2} m_1(A_1) \cdot m_2(A_2).$$

Renormalisation is then of no use, since this disjunctive fusion is a union of random sets, but the result is more imprecise. The resulting belief function Bel is the product of Bel_1 and Bel_2: $Bel(A) = Bel_1(A) \cdot Bel_2(A), \forall A \subseteq S$. Applied to probability distributions p_1 and p_2, the result of the disjunctive fusion is no longer a probability measure but a belief function whose focal sets are singletons or 2-elements subsets (the closure property is violated).

Rather than adopting a disjunctive combination uniformly, alternative options have been proposed for handling the mass of conflict $CONF = \sum_{A_1,A_2:A_1\cap A_2=\emptyset} m_1(A_1) \cdot m_2(A_2)$ when it is too large:

- Smets[112] suggested to abstain from renormalizing, thus laying bare the conflict. Then the unnormalized Dempster rule comes down to multiplying commonality functions $Q_i(A) = \sum_{A\subseteq E} m_i(E)$.
- Yager [125] proposed to assign it to the whole frame S, turning inconsistency into ignorance.

- Dubois and Prade [44] allocate the mass $m_1(A_1) \cdot m_2(A_2)$ to the set $A_1 \cap A_2$ if it is not empty, and to the disjunction $A_1 \cup A_2$ otherwise.
- Other authors share the mass $m_1(A_1) \cdot m_2(A_2)$ between A_1 and A_2 when they are disjoint [108], and more generally between $A_1 \cap A_2$, $A_2 \setminus A_1$, and $A_1 \setminus A_2$ regardless of whether A_1 and A_2 are disjoint or not [127].

Under such schemes, associativity is generally lost. An extensive comparative discussion of fusion rules in the theory of evidence is provided by Smets [115].

Belief functions are also compatible with a combination mode based on weighted average. Indeed, the weighted arithmetic mean of mass functions is a mass function. The belief function $Bel = \sum_{i=1}^{n} \alpha_i \cdot Bel_i$ has a mass function $\sum_{i=1}^{n} \alpha_i \cdot m_i$. This is a generalisation of non-Bayesian probabilistic fusion, which also applies to the fusion of possibility measures, however without preserving the nestedness of focal sets. So, the weighted arithmetic mean of products of belief functions is a belief function. The arithmetic mean is instrumental for the discounting of a belief function provided by a source with low reliability as pointed out by Shafer [104]. Let α be the probability that the source providing the belief function Bel_1 is reliable. It means that, with a probability $1 - \alpha$, nothing is known, which corresponds to a second source providing the non-informative mass function $m_2(S) = 1$. The weighted arithmetic mean of these mass functions is $m = \alpha \cdot m_1 + (1 - \alpha) \cdot m_2$. The mass allocated to the informative subset $A \subset S$ decreases since $m(A) = \alpha \cdot m_1(A)$, while the mass allocated to the whole frame, i.e. the tautology $(m(S) = \alpha \cdot m_1(S) + (1 - \alpha))$ increases.

It is not very easy to find natural idempotent *conjunctive* fusion rules for belief functions using mass functions. Dubois and Yager [58] propose a methodology for building such fusion rules by duplicating focal sets and sharing masses so as to make the two mass functions commensurate. However there is no unique combination scheme resulting from this process, even if this approach enables the minimum rule to be retrieved if the two belief functions are consonant[27]. Dubois et al. [55] show that the minimum rule of possibility theory can be interpreted as a minimal commitment fusion of consonant belief functions, in the sense of the commonality-based information ordering. Recently Denoeux [26] proposes to use the decomposition of a non-dogmatic belief function ($m(S) > 0$) as a Dempster combination of simple support functions,

$$m = \oplus_{A \subset S} A^{w(A)}$$

where $A^{w(A)}$ denotes the simple support belief function with mass function m_A such that $m_A(A) = 1 - w(A)$ and $m(S) = w(A)$. In fact, not all belief functions can be expressed this way, unless we admit that some terms in the above equation are fictitious simple support belief functions for which $w(A) > 1$. Then the decomposition exists and is unique for non-dogmatic belief functions[112]. The idea of the idempotent rule is then to combine weight functions w event-wise using the minimum. However, when applied to consonant belief functions, it does not retrieve the minimum rule of possibility theory. The question of finding a canonical idempotent fusion rule in the theory of evidence consistently with the one in possibility theory is still an open problem.

7.5 Merging imprecise probability families

The fusion of imprecise probabilities is not really in agreement with the fusion of belief functions. Given two familles $\mathcal{P}_1$ and $\mathcal{P}_2$ of probabilities provided by two reliable sources, it is natural to consider that the result of a fusion is the intersection $\mathcal{P} = \mathcal{P}_1 \cap \mathcal{P}_2$, when non-empty. In contrast with Dempster rule, this fusion mode is idempotent. But, while it sounds hard to justify Dempster fusion rule in terms of imprecise probabilities, in the same way it is not easy to express the mass function induced by $\mathcal{P}_1 \cap \mathcal{P}_2$ in terms of the mass functions induced by $\mathcal{P}_1$ and $\mathcal{P}_2$. One may apply the idempotent fusion of imprecise probabilities to belief functions, by performing the intersection of sets $\mathcal{P}_i = \{P : P(A) \geq Bel_i(A), \forall A \subseteq S\}$ for $i = 1, 2$. But the lower bounds of the induced probabilities are not belief functions in general. Chateauneuf [11] explores these issues in some detail. However many questions remain open.

8 Conclusion

This chapter offers an overview of uncertainty representation frameworks where the problems of collecting observations tainted with variability and of representing incomplete information are carefully distinguished. The first one naturally leads to a probabilistic approach, while the second situation is more naturally described in terms of sets of mutually exclusive elements, and belongs to the realm of logic (if the variables describing a problem are Boolean), or of interval analysis (for numerical variables). The existing new uncertainty theories are hybrids of these basic approaches, some variants being purely ordinal. It including the case of linguistic information dealing with numerical variables (fuzzy set theory). This synergy between uncertainty representation frameworks is fruitful since it provides very expressive formal tools for the faithful representation of pieces of information along with their imperfections. It contrasts with the Bayesian theory of subjective probabilities, which looks at a loss to ensure a clear distinction between uncertainty due variability and uncertainty due to ignorance. The unified view offered here also enables formal notions coming from set theory or probability theory to be generalised to other settings. For example, one can introduce conditioning, independently from the notion of probability, even in symbolic representations of the logical type, or use logical connectives for combining probabilities, fuzzy sets, random sets, etc. Lastly, injecting interval analysis into the notion of mathematical expectation gets close to non-additive Choquet integrals studied in other chapters, for belief functions, or possibility measures.

Acknowledgements The authors thank Claudette Cayrol for her useful comments on the French initial version of this text.

References

[1] E.W. Adams and H.P. Levine. On the uncertainties transmitted from premises to conclusions in deductive inferences. *Synthese*, 30:429–460, 1975.

[2] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51(1):86–108, 2006.

[3] S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92:259–276, 1997.

[4] J.-M. Bernard. An introduction to the imprecise dirichlet model for multinomial data. *Int. J. Approx. Reasoning*, 39(2-3):123–150, 2005.

[5] Z. W. Birnbaum. On random variables with comparable peakedness. *Annals of Mathematical Statistics*, 19:76–81, 1948.

[6] G. Boole. *An Investigation of the Laws of Thought on which are Founded the Mathematical Theory of Logic and Probabilities.* MacMillan, reprinted by Dover, New York, 1958.

[7] D. Budescu and T. Wallstein. Processing linguistic probabilities: general principles and empirical evidence. In J.R. Busemeyer, R. Hastie, and D. Medin, editors, *The Psychology of Learning and Motivation: Decision-Making from the Perspective of Cognitive Psychology*, pages 275–318. Academic Press, 1995.

[8] P. Calabrese. An algebraic synthesis of the foundations of logic and probability. *Information Sciences*, 42:187–237, 1987.

[9] T. Calvo, G. Mayor, and R. Mesiar, editors. *Aggregation Operators: New Trends and Applications*, volume 97 of *Studies in Fuzziness and Soft Computing.* Physica-Verlag, Heidelberg, 2002.

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

[11] A. Chateauneuf. Combination of compatible belief functions and relation of specificity. In M. Fedrizzi, J. Kacprzyk, and R.R. Yager, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 98–114. Wiley, New York, 1994.

[12] A. Chateauneuf. Decomposable capacities, distorted probabilities and concave capacities. *Math. Soc. Sciences*, 31:19–37, 1996.

[13] A. Chateauneuf and J.Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of moebius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.

[14] B.F. Chellas. *Modal Logic.* Cambridge University Press, Cambridge, UK, 1978.

[15] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, 5(4):131–295, 1953.

[16] G. Coletti and R. Scozzafava. *Probabilistic Logic in a Coherent Setting.* Kluwer Academic Pub, 2002.

[17] G. Coletti and R. Scozzafava. Coherent conditional probability as a measure of uncertainty of the relevant conditioning events. In *Proc. of ECSQARU03*, LNAI 2711, pages 407–418, Aalborg, 2003. Springer Verlag.

[18] R. M. Cooke. *Experts in Uncertainty.* Oxford University Press, Oxford, UK, 1991.

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

[20] G. de Cooman. A behavioural model for vague probability assessments. *Fuzzy Sets and Systems*, 154(3):305–358, 2005.

[21] G. De Cooman and D. Aeyels. Supremum-preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.

[22] B. De Finetti. La logique de la probabilité. In *Actes du Congrès Inter. de Philosophie Scientifique (1935)*, pages IV1–IV9, Paris, France, 1936. Hermann et Cie Editions.

[23] B. De Finetti. La prévision : ses lois logiques, ses sources subjectives. *Ann. Inst. Poincaré*, 7:1–68, 1937.

[24] B. De Finetti. *Theory of probability.* Wiley, New York, 1974.

[25] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.*, 38:325–339, 1967.

[26] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artif. Intell.*, 172(2-3):234–264, 2008.

[27] S. Destercke, D. Dubois, and E. Chojnacki. Cautious conjunctive merging of belief functions. In Khaled Mellouli, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, volume 4724 of *Lecture Notes in Artificial Intelligence, LNAI*, pages 332–343, http://www.springerlink.com, 2007. Springer.

[28] S. Destercke, D. Dubois, and E. Chojnacki. Possibilistic information fusion using maximal coherent subsets. In *IEEE International Conference on Fuzzy Systems, London (UK)*, pages 445–450, http://www.ieee.org/, 2007. IEEE.

[29] S. Destercke, D. Dubois, and E. Chojnacki. Relating practical representations of imprecise probabilities. In G. De Cooman, J. Vejnarova, and M. Zaffalon, editors, *International Symposium on Imprecise Probability: Theories and Applications (ISIPTA), Prague (Czech Republic)*, pages 155–164, http://www.sipta.org, 2007. SIPTA.

[30] Z. Domotor. Probability kinematics - conditional and entropy principles. *Synthese*, 63:74–115, 1985.

[31] D. Dubois, , and E. Huellermeier. Comparing probability measures using possibility theory: A notion of relative peakedness. *International Journal of Approximate Reasoning*, 45:364–385, 2007.

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

[33] D. Dubois, H. Fargier, and H. Prade. Possibilistic likelihood relations. In *Proceedings of 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'98)*, pages 1196–1202, Paris, France, Jul 1998. Editions EDK.

[34] D. Dubois, H. Fargier, and H. Prade. Ordinal and probabilistic representations of acceptance. *J. Artif. Intell. Res. (JAIR)*, 22:23–56, 2004.

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

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

[37] D. Dubois, S. Moral, and H. Prade. Belief change rules in ordinal and numerical uncertainty theories. In D. Dubois and H. Prade, editors, *Belief Change*, pages 311–392. Kluwer Acad. Publ., 1998.

[38] D. Dubois, H. T. Nguyen, and H. Prade. Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy Sets*, The Handbooks of Fuzzy Sets Series, pages 343–438. Kluwer, Boston, Mass., 2000.

[39] D. Dubois and H. Prade. 0n several representations of an uncertain body of evidence. In M.M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, Amsterdam, 1982.

[40] D. Dubois and H. Prade. Unfair coins and necessity measures: a possibilistic interpretation of histograms. *Fuzzy Sets and Systems*, 10(1):15–20, 1983.

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

[42] D. Dubois and H. Prade. Properties of information measures in evidence and possibility theories. *Fuzzy Sets and Systems*, 24:161–182, 1987.

[43] D. Dubois and H. Prade. *Théorie des Possibilités - Applications à la Repréâsentation des Connaissances en Informatique.* Masson, Paris, 1987.

[44] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(4):244–264, 1988.

[45] D. Dubois and H. Prade. Aggregation of possibility masures. In J. Kacprzyk and M. Fedrizzi, editors, *Multiperson Decision Making using Fuzzy Sets and Possibility Theory*, pages 55–63. Kluwer, Dordrecht, 1990.

[46] D. Dubois and H. Prade. Consonant approximations of belief functions. *Int. J. Approximate Reasoning*, 4(5/6):419–449, 1990.

[47] D. Dubois and H. Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, 42:87–101, 1991.

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

[49] D. Dubois and H. Prade. Conditional objects as nonmonotonic consequence relationships. *Special issue on Conditional Event Algebra, IEEE Trans. on Systems, Man and Cybernetics*, 24(12):1724–1740, 1994.

[50] D. Dubois and H. Prade. La fusion d'informations imprécises. *Traitement du Signal*, 11(6):447–458, 1995.

[51] D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In P. Smets, editor, *Handbook on Defeasible Reasoning and Uncertainty Management Systems - Volume 1: Quantified Representation of Uncertainty and Imprecision*, pages 169–226. Kluwer Academic Publ., Dordrecht, The Netherlands, 1998.

[52] D. Dubois and H. Prade. Possibility theory in information fusion. In G. Della Riccia, H. Lenz, and R. Kruse, editors, *Data Fusion and Perception*, Vol. 431 in the CISM Courses and Lectures, pages 53–76. Springer-Verlag, Berlin, 2001.

[53] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In R. Lowen and M. Roubens, editors, *Fuzzy Logic. State of the Art*, pages 103–112. Kluwer Acad. Publ., Dordrecht, 1993.

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

[55] D. Dubois, H. Prade, and P. Smets. New semantics for quantitative possibility theory. In *ISIPTA '01, Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications, Ithaca, NY, USA*, pages 152–161. Shaker, 2001.

[56] D. Dubois, H. Prade, and P. Smets. A definition of subjective possibility. *Badania Pperacyjne i Decyzje (Operation research ans decisions)*, 4:7–22, 2003.

[57] D. Dubois and A. Ramer. Extremal properties of belief measures in the theory of evidence. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 1(1):57–68, 1993.

[58] D. Dubois and R. Yager. Fuzzy set connectives as combinations of belief structures. *Information Sciences*, 66:245–275, 1992.

[59] S. Ferson, L. Ginzburg, V. Kreinovich, D.M. Myers, and K. Sentz. Construction probability boxes and Dempster-Shafer structures. Technical report, Sandia National Laboratories, 2003.

[60] S. Ferson and L. R. Ginzburg. Different methods are needed to propagate ignorance and variability. *Reliability Engineering and System Safety*, 54:133–144, 1996.

[61] T. Fine. *Theories of Probability.* Academic Press, New-York, 1983.

[62] J. Fodor and R. Yager. Fuzzy set-theoretic operators and quantifiers. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy Sets*, The Handbook of Fuzzy Sets Series, pages 125–193. Kluwer Academic Publ., Dordrecht, 2000.

[63] S. French. Group consensus probability disributions: a critical survey. In J. subjectivist frameworko and et al., editors, *Bayesian Statistics 2*, pages 183–202. Elsevier Science, 1985.

[64] N. Friedman and J. Halpern. Plausibility measures and default reasoning. In *Proc of the 13th National Conf. on Artificial Intelligence*, pages 1297–1304, Portland, OR, 1996.

[65] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, Cambridge, Mass, 1988.

[66] I. Gilboa and D. Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

[67] I. Gilboa and D. Schmeidler. Updating ambiguous beliefs. In Y. Moses, editor, *Theoretical Aspects of Reasoning About Knowledge (Proc. of the 4th Conf. TARK'92)*, pages 143–162, 1992.

[68] Itzhak Gilboa and David Schmeidler. Case-based knowledge and induction. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 30(2):85–95, 2000.

[69] R. Giles. Foundations for a theory of possibility. In M.M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 1836–195. North-Holland, 1982.

[70] I.J. Good. Subjective probability as the measure of a non-measurable set. In E. Nagel and et al., editors, *Logic, Probability and Philosophy of Science*. Stanford University Press, Stanford, 1962.

[71] I.R. Goodman, H.T. Nguyen, and E.A. Walker. *Conditional Inference and Logic for Intelligent Systems: A Theory of Measure-Free Conditioning*. North-Holland, Amsterdam, 1991.

[72] I. Hacking. *The Emergence of Probability*. Cambridge University Press, Cambridge, UK, 1975.

[73] J. Halpern. Defining relative likelihood in partially-ordered preferential structures. *J. AI Research*, 7:1–24, 1997.

[74] J. Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, N.Y., 1962.

[75] J.Y. Jaffray. Bayesian updating and belief functions. *IEEE Trans. on Systems, Man and Cybernetics*, 22:1144–1152, 1992.

[76] E.T. Jaynes. Where do we stand on maximum entropy. In I. Levine and M. Tribus, editors, *The Maximum Entropy Formalism*. MIT Press, Cambridge USA, 1979.

[77] D. Kahnemann, P. Slovic, and A. Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, UK, 1979.

[78] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Pub., Boston, 2000.

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

[80] B.O. Koopman. The bases of probability. *Bull. Am. Math. Soc.*, 46:763–774, 1940.

[81] C.H. Kraft, J.W. Pratt, and A. Seidenberg. Intuitive probability on finite sets. *Ann. Math. Stat.*, 30:408–419, 1959.

[82] E. Kriegler and H. Held. Utilizing belief functions for the estimation of future climate change. *I. J. of Approximate Reasoning*, 39:185–209, 2005.

[83] H. Kyburg. *The Logical Foundations of Statistical Inference*. Reidel, Dordrecht, the Netherlands, 1974.

[84] K. Lehrer and C. Wagner. *Rational Consensus in Science and Society*. Reidel, Dordrecht, 1981.

[85] I. Levi. *The Enterprize of Knowledge*. The MIT Press, Cambridge, Mass., 1980.

[86] D.K. Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85:297–315, 1976.

[87] D.K. Lewis. *Counterfactuals*. Basil Blackwell (1973), 2nd edition, Billing and Sons Ltd., Worcester, UK, 1986.

[88] D.V. Lindley. Scoring rules and the inevitability of probability. *Int. Statist. Rev.*, 50:1–26, 1982.

[89] R.D. Luce and H.P. Raiffa. *Games and Decisions*. Wiley, New York, 1981.

[90] J. Marschak. Personal probabilities of probabilities. *Theory and Decision*, 6:121–125, 1975.

[91] E. Miranda, I. Couso, and P. Gil. Relationships between possibility measures and nested random sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(1):1–15, 2002.

[92] E. Miranda, I. Couso, and P. Gil. A random set characterization of possibility measures. *Inf. Sci.*, 168(1-4):51–75, 2004.

[93] R. E. Moore. *Interval analysis*. Prentice-Hall, Englewood Cliffs, N.J., 1966.

[94] D. Mundici. The logic of ulam games with lies. In C. Bicchieri and M. Dalla Chiara, editors, *Knowledge*, Belief and Strategic Interaction, pages 275–284. Cambridge Univ. Press, 1992.

[95] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.

[96] J. Paris. *The uncertain Reasoner's Companion*. Cambridge University Press, 1994.

[97] Z. Pawlak. *Rough Sets - Theoretical Aspects of Reasning about Data*. Kluwer Academic Publ., Dordrecht, 1991.

[98] F. P. Ramsey. Truth and probability. In J. Kyburg and H.E. Smokler, editors, *Studies in Subjective Probability*, pages 23–52. Krieger Pub. Co, Huntington, N.Y., 1980. originalement publi en 1926.

[99] E. Raufaste, R. da Silva Neves, and C. Mariné. Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence*, 149:197–218, 2003.

[100] N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, 1:179–219, 1970.

[101] L.J. Savage. *The Foundations of Statistics*. Wiley (1954); 2nd edition, Dover Publications Inc., New York, 1972.

[102] D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57:571–587, 1989.

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

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

[105] G. Shafer. Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences*, 19:309–370, 1978.

[106] G. Shafer. Allocations of probability. *Ann. Probab*, 7:827–839, 1979.

[107] L. S. Shapley. A value for n-person games. In Kuhn and Tucker, editors, *Contributions to the Theory of Games, II*, pages 307–317. Princeton University Press, 1953.

[108] F. Smarandache and J. Dezert. Proportional conflict redistribution rules for information fusion. In Dezert J. Smarandache F., editor, *Applications and Advances of DSmT for Information Fusion. Vol. 2.*, pages 3–68. American Research Press, Rehoboth, N.M., 2006.

[109] P. Smets. Possibilistic inference from statistical data. In *Proc. of the 2nd World Conf. on Mathematics at the Service of Man*, pages 611–613, Las Palmas (Canary Island), Spain, Jun-Jul 1982.

[110] P. Smets. Constructing the pignistic probability function in a context of uncertainty. In M. Henrion and et al., editors, *Uncertainty in Artificial Intelligence*, pages 29–39. North-Holland, Amsterdam, 1990.

[111] P. Smets. The normative representation of quantified beliefs by belief functions. *Artificial Intelligence*, 92:229–242, 1997.

[112] P. Smets. The transferable belief model for quantified belief representation. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 267–301. Kluwer Academic, Dordrecht, Pays-Bas, 1998.

[113] P. Smets. Quantified possibility theory seen as an hypercautious transferable belief model. In *Rencontres Francophones sur les Logiques Floues et ses Applications (LFA 2000)*, pages 343–353, La Rochelle, France, Oct 2000. Cepadues-Editions.

[114] P. Smets. Belief functions on real numbers. *nternational Journal of Approximate Reasoning*, 40:181–223, 2005.

[115] P. Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4):387–412, 2007.

[116] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

[117] C.A.B. Smith. Consistency in statistical inference and decision. *J. Royal Statis. Soc.*, B-23:1–37, 1961.

[118] T.H. Strat. Continuous belief functions for evidential reasoning. In *National Conference on Artificial Intelligence (AAAI-84)*, pages 308–313, Los Altos, CA, 1984. W. Kaufmann.

[119] M. Sugeno. Fuzzy measures and fuzzy integral : a survey. In M.M. Gupta, G.N. Saridis, and B.R. Gaines, editors, *Fuzzy Automata and Decision Processes*, pages 89–102. North-Holland, Amsterdam, 1977.

[120] P. Teller. Conditionalization, observation and change of preference. In W.L. Harper and C.H. Hooker, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume 1, pages 205–259. D. Reide, Los Altos, CA, 1984.

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

[122] P. Walley and T. Fine. Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, 10:741–761, 1982.

[123] P.M. Williams. Bayesian conditionalization and the principle of minimum information. *British J. for the Philosophy of Sciences*, 31:131–144, 1980.

[124] S.M.K. Wong, Y.Y. Yao, P. Bollman, and H.C. Burger. Axiomatization of qualitative belief structure. *IEEE trans. Systems Man and Cybern*, 21:726–734, 1993.

[125] R. R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41(2):93–137, 1987.

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

[127] K. Yamada. A new combination of evidence based on compromise. *Fuzzy Sets and Systems*, 159(13):1689–1708, 2008.

[128] L. A. Zadeh. Fuzzy probabilities. *Inf. Process. Manage.*, 20(3):363–372, 1984.

[129] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[130] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning, part i. *Information Sciences*, 8:199–249, 1975.

[131] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning, part ii. *Information Sciences*, 8:301–357, 1975.

[132] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning, part iii. *Information Sciences*, 9:43–80, 1975.

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

[134] L.A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, 90:111–127, 1997.