Open-frame Dempster Conditioning for Incomplete Interval Probabilities

Paola Castellan
DSM, University, Trieste, I-34100

Andrea Sgarro*
DSM, University, Trieste, I-34100
sgarro@univ.trieste.it

Abstract

The second author has put forward a theory of incomplete interval probabilities meant to give a common framework to both interval probabilities and open-frame bodies of evidence, as obtained by application of the non-normalized (open-frame) Dempster rule. Below we re-describe this proposal and then compare two possible ways of "conditioning" based on the open-frame Dempster rule: namely, we condition the original (possibly incomplete) knowledge by pooling it with new evidence which assigns certainty to the conditioning event. The idea is trying to build a probabilistic theory which would be able to cope not only with uncertainty and ignorance, but also with [forms of] contradictoriness, to be included into the description of a possible state of knowledge.

Keywords. Open-frame beliefs, Dempster conditioning, incomplete probabilities, interval probabilities.

1 Introduction

In [3] the second author has put forward a theory of incomplete interval probabilities meant to give a common framework to both usual (complete) interval probabilities and open-frame bodies of evidence, as obtained by application of the non-normalized (open-frame) Dempster rule. Below we re-describe this proposal in section 3; then, in section 4, we compare two possible ways of "conditioning" based on the open-frame Dempster rule: namely, we condition the original (possibly incomplete) knowledge by pooling it with new evidence which assigns certainty to the "conditioning" event. To avoid confusion, we shall reserve the name "open-frame Dempster conditioning" only to the second way, and call "slicing" the first way (cf section 4). In Section 5 we comment on the purport of our proposal.

The proposal in [3] has been inspired by three (comparatively) old ingredients: interval probabilities, open-frame bodies of evidence and Rényi's incomplete probabilities (in the sequel "open-frame" and "incomplete" are interchangeable; the same for "closed-frame" and "complete"). In the second ingredient, i.e. open-frame bodies of evidence as obtained by use of the open-frame Dempster rule, a formal novelty is introduced, which has also a conceptual bearing: actually, it is precisely this formal change that lead us to obtain a unifying approach to available theories, and this more or less "at no mathematical cost". The unification was obtained by putting forward an interval-type theory of incomplete probabilities.

We defer philosophical comments meant to vindicate our proposal to the final section 5, after the proposal has been described and a few examples have been given. However, we wish to make soon clear a few points. If our proposal is taken as a merely *formal* one, motivated only by mathematical convenience, we think that it needs no special justification. If instead one wants to understand whether our formalism has also some proper contents, then one has to be more wary. We wish to stress soon that in this case one has to accept at least the following three facts, as sort of "working hypotheses", else the proposal falls. First we are taking for granted the adequacy of the open-world variant of Dempster rule for pooling opinions, and so obtaining new, possibly incomplete, states of knowledge. More precisely (and less assumingly), we need only a very special case of Dempster rule, i.e. Dempster conditioning without the normalization coefficient, as defended by Smets in his seminal paper [6] (cf section 3 and 4). Dempster rule has been the object of an extensive debate in the literature, cf e.g.[9] or [6]; if it is rejected (or rather: if even the special case we need is rejected) our proposal becomes largely irrelevant. It is by Dempster-pooling two *dissonant* states of knowledge that one obtains a new *self-dissonant* state of knowledge: the latter is an

*Partially supported by MURST and CNR.

open-frame body of evidence [1] and can be described by means of a family of incomplete probability vectors, which are a sort of epistemic counterparts to Rényi's incomplete probabilities (cf below subsection 3.4). These incomplete probability vectors are the "building-blocks" of our proposal. The second working hypothesis is that the formal inclusion between closed-frame evidence and interval probabilities has a purport also at a philosophical level; if one finds this inclusion irrelevant, save on the formal plane, one will hardly be ready to accept an inclusion such as ours, which is even more assuming (cf subsection 3.4). The third working hypothesis appears to be less assuming, but we have to mention it explicitly all the same. Namely, we assume that a state of knowledge is not modified when one adds what Smets calls in [6] impossible propositions, that is if one enlarges the frame of discernment by adding objects which have zero confidence, be this confidence a probability, a belief or a plausibility (cf below subsection 3.2).

2 Notational Preliminaries

Our "universe", or *frame of discernment*, is a finite set $\mathcal{X} = \{a_1, a_2, \ldots, a_K\}$ of $K \geq 1$ elements, or "elementary events" (generalizations to the countable and the continuous case are feasible). We shall deal with *confidence distributions*, called also *fuzzy measures*, or *Choquet capacities*, i.e.monotonic set-functions Φ constrained to be non-negative and upper-bounded by unity:

$$\Phi(\emptyset) = 0, \;\; \Phi(\mathcal{X}) \leq 1,$$
$$A \subseteq B \;\Rightarrow\; \Phi(A) \leq \Phi(B) \qquad (1)$$

We list some additional properties Φ may have (the arrow denotes implication). If $\Phi(\mathcal{X}) = 1$, Φ is *complete* (*normal[ized], regular*). If $A \cap B = \emptyset \Rightarrow \Phi(A \cup B) \leq \Phi(A) + \Phi(B)$, Φ is *subadditive*, while Φ is *superadditive* if $A \cap B = \emptyset \Rightarrow \Phi(A \cup B) \geq \Phi(A) + \Phi(B)$. If $\Phi(A \cup B) + \Phi(A \cap B) \leq \Phi(A) + \Phi(B)$ whatever A and B, Φ is *strongly subadditive*, while Φ is *strongly superadditive* if $\Phi(A \cup B) + \Phi(A \cap B) \geq \Phi(A) + \Phi(B)$ $(A, B \subseteq \mathcal{X})$.

Actually we shall use *couples* $\{\Phi_*,\ \Phi^*\}$ of set-functions as in (1), where Φ_* is *dominated* by Φ^*, i.e. $\Phi_*(A) \leq \Phi^*(A)$, $\forall A \subseteq \mathcal{X}$; Φ_* is the *lower* confidence, while Φ^* is *upper* confidence; to each subset A we can associate the interval $[\Phi_*(A),\ \Phi^*(A)]$. Needless to say, a single set-function Φ can be viewed as an *interval* set-function $\{\Phi_*,\ \Phi^*\}$ by letting all the intervals boil down to points: $\Phi_*(A) = \Phi^*(A) = \Phi(A)$. With more generality (*iff* stands for *if and only if*):

Definition. $\{\Phi_*, \Phi^*\}$ is *flat* over A iff $\Phi_*(A) = \Phi^*(A)$.

We shall omit mentioning the set, when flatness is over the entire frame $\mathcal{X}$.

3 The Three Ingredients and the New Recipe

3.1 The First Ingredient: Complete Interval Probabilities

(Cf e.g. [7].) Let $\mathcal{P}$ be a non-empty family of probability vectors $P = (p_1, p_2, \ldots, p_K)$ over the universe $\mathcal{X}$ ($p_i \geq 0$, $\sum_i p_i = 1$); without real restriction one can assume that $\mathcal{P}$ is convex and topologically closed. One defines the *lower* and the *upper probability* of A, $P_*(A)$ and $P^*(A)$, respectively, as

$$P_*(A) = \min_{P: P \in \mathcal{P}} P(A)\ , \quad P^*(A) = \max_{P: P \in \mathcal{P}} P(A) \qquad (2)$$

3.2 The Second Ingredient: Open-frame Bodies of Evidence

(Cf e.g. [4] and [6].) We go to a subtheory of interval probabilities called *evidence theory*[2]. One assigns a (formal!) probability vector m over $2^{\mathcal{X}} - \emptyset$, rather than over $\mathcal{X}$. The meaning of the number $m(A)$ is the *weight of evidence* the expert has in support of A, while she is ignorant about how to divide this support among strict subsets of A. If $m(F) \neq 0$, F is called a *focal set*. The weight m defines a *body of evidence* over $\mathcal{X}$. Two more numbers are associated with A; the *belief* of A cumulates all the evidence that directly supports A, while the *plausibility* of A cumulates all the evidence that does not directly oppose A:

$$\mathrm{Bel}(A) \;=\; \sum_{F: F \subseteq A} m(F)\ ,$$

$$\mathrm{Pl}(A) \;=\; \sum_{F: F \cap A \neq \emptyset} m(F) = 1 - \mathrm{Bel}(\overline{A}) \qquad (3)$$

Some see the interval $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$ as the *interval probability* of A, its *point probability* remaining unspecified (actually, in a way which is both easy and natural, one can construct a family $\mathcal{P}$ such that

[1] So incompleteness signals "self-dissonance", or "self-conflict". The expression "self-contradictory" is also used in the literature; this gives a pejorative connotation which is harsh indeed. In a way, if interval-type theories add "ignorance" to "uncertainty" as covered by Bayesian pointwise probabilities, incomplete theories allow one to describe also [forms of] contradictoriness.

[2] It is a moot point whether the inclusion between the two theories is only formal, or also conceptual. More or less everybody, however, agrees that a Bayesian probability vector and the corresponding Bayesian body of evidence describe the same state of knowledge.

$\mathrm{Bel} = P_*$ and $\mathrm{Pl} = P^*$; cf subsection 3.4). All the intervals boil down to points when the body of evidence is *Bayesian*, i.e. when all the focal sets are singletons.

In evidence theory, to pool two bodies of evidence, m_1 and m_2, into a new body of evidence m one uses *Dempster rule*, which is the chief *inferential engine* of the theory:

$$m(A) \propto \sum_{F,G:F\cap G=A} m_1(F)\times m_2(G), \quad A\neq\emptyset \quad (4)$$

If any focal set of the first body of evidence intersects any focal set of the second, the proportionality sign $\propto$ in the formula simply becomes an equality sign $=$. Unfortunately, when the two bodies are instead *dissonant*, some difficulties, also conceptual, are found, so far as one insists on having $m(\emptyset)=0$. A bold way-out has been taken in the literature by dropping this requirement and allowing

$$m(\emptyset) = \sum_{F,G:F\cap G=\emptyset} m_1(F)\times m_2(G) > 0$$

The enlarged theory is an *open-frame* theory, which allows for an open frame of discernment $\mathcal{X}$ (*open-frame* assumption $\sum_{F:\ F\neq\emptyset} m(F)\leq 1$ as opposed to the *closed-frame* assumption $\sum_{F:\ F\neq\emptyset} m(F) = 1$; cf [6][3]). This way, however, the interpretation of beliefs and plausibilities as lower and upper probabilities is lost, or so it would appear (cf instead subsection 3.4).

In open-frame evidence theory beliefs are defined as in (2), only ruling out the (possible) focal set $\emptyset$ from the first summation, which is taken over $\{F:\emptyset\neq F\subseteq A\}$. Following [3], we shall adopt a slightly different, and we think more convenient, approach: we demand $m(\emptyset)=0$, just as in the closed-frame theory, but allow $\sum_F m(F)\leq 1$. In other words, the weight $1-\sum_{F:\ F\neq\emptyset} m(F)$ needed to reach unity (the "degree of incompleteness" of the frame) is left lacking, rather than being given to the empty set. We stress that the difference between $m(\emptyset)=0$, $\sum_F m(F)\leq 1$ on one side and $m(\emptyset)\geq 0$, $\sum_F m(F)=1$ on the other is more formal than conceptual, since $m(\emptyset)$ does not concur in forming beliefs and plausibilities, anyway. So we are pleading for a formal change, such as to make life easier; however, as is often the case, this formal change suggests a conceptual change.

To sum up: in our setting an *incomplete body of evidence* (an *open-frame body of evidence*), will be described through an *incomplete weight* m:

$$m(F)\geq 0\ ,\ \ m(\emptyset)=0\ ,\ \ \sum_F m(F)\leq 1$$

[3]In [6] the rule is defended by putting forward a set of axioms, or "desirable properties", which an adequate rule for pooling opinions should verify, and which are verified only by rule (4).

Incomplete beliefs and plausibility are computed exactly as in (3), and are interpreted in the same way as there:

$$\mathrm{Bel}(A) = \sum_{F:F\subseteq A} m(F)\ ,$$

$$\mathrm{Pl}(A) = \sum_{F:F\cap A\neq\emptyset} m(F) \quad (5)$$

If the reader is ready to accept the following statement: two bodies of evidence over $\mathcal{X}$ and $\mathcal{X}\cup\{z\}$ ($z\notin\mathcal{X}$) coincide when they are described by the same focal sets with the same weights (z cannot belong to any focal set), then it is straightforward to prove that *any* incomplete body of evidence m over $\mathcal{X}$ can be obtained by Dempster-pooling two complete ones, m_1 and m_2, say; to see this, just define m_1 by setting $m_1(z) = 1-\sum_F m(F)$, else $m_1 = m$, and define m_2 by setting $m_2(\mathcal{X})=1$. In particular, if m_1 is a Bayesian body of evidence, one obtains an incomplete state of knowledge m of a type which is the base for the subsequent discussion.

Note that we have just used the rule in a very special case, i.e. when one of the two bodies of evidence to be pooled is *unifocal* (and complete): $\exists C:\ m_2(C)=1$. This is the *only* case when we need the Dempster rule, and so we shall be more specific; in [6] it is argued that this special form is actually a more basic and fundamental principle than the rule as a whole. Below m_1 is a possibly incomplete body of evidence to be pooled with m_2, $m_2(C)=1$; $m=m_1\otimes m_2$ is the resulting weight after pooling m_1 and m_2; $\mathrm{Bel}^{(1)}$ and $\mathrm{Pl}^{(1)}$ refer to beliefs and plausibilities with respect to m_1, that is *before* pooling, while Bel and Pl refer to beliefs and plausibilities *after* pooling. We assume to no restriction $\emptyset\neq A\subseteq C$ (else $m(A)=0$, $\mathrm{Bel}(A)=\mathrm{Bel}(A\cap C)$ and $\mathrm{Pl}(A)=\mathrm{Pl}(A\cap C)$). One has from (4):

$$m(A) = \sum_{F:\ F\cap C=A} m_1(F)\ ,$$

and so (5) becomes

$$\mathrm{Bel}(A) = \mathrm{Bel}^{(1)}(A\cup\overline{C}) - \mathrm{Bel}^{(1)}(\overline{C})\ ,$$

$$\mathrm{Pl}(A) = \mathrm{Pl}^{(1)}(A\cap C) \quad (6)$$

These identities are soon checked; the expression for $\mathrm{Bel}(A)$ can be obtained by first writing it as a sum $\sum m_1(F)$ taken over the focal sets F such that $\emptyset\neq F\cap C\subseteq A$. In Section IV the equalities in (6) will be extended and re-interpreted as a form of conditioning, to be compared with yet another form of conditioning called "slicing".

3.3 The Third Ingredient: Incomplete Probabilities

(Cf e.g. [2].) It is not generally known that incomplete frames of discernment have a long-standing tradition in probability theory: *incomplete* probability vectors P for which $\sum_i p_i \leq 1$ have been studied by no less than A. Rényi; actually Rényi had in mind an *objective* view of probability, based on observed frequencies, rather than an *epistemic* view, based on degrees of confidence, such as the one we are pursuing here, and so in his vision empirical non-observability takes the place of epistemic self-dissonance. Non-negativity for the components p_i of vector P is of course assumed; P is additive:

$$P(A) = \sum_A p_i \equiv \sum_{a_i \in A} p_i$$

and so $P(\emptyset) = 0$. We shall call $P(\mathcal{X}) = \sum_i p_i$ the *load* of the probability vector.

We observe that an incomplete probability vector P *à la Rényi* can be accomodated into evidence theory in *two* ways: in a closed-frame setting one can set $m(a_i) = p_i$, $m(\mathcal{X}) = 1 - P(\mathcal{X})$, while in an open-frame setting one can set $m(a_i) = p_i$, $m(\mathcal{X}) = 0$, which preserves the additivity of the monotonic set-function $P = \text{Bel} = \text{Pl}$. In these pages the default choice is the second, i.e. the open-frame one with $m(\mathcal{X}) = 0$. To our view the first way appears to be even less "philological" than the second, since Rényi had clearly in mind an open-frame setting (cf examples 1 and 2 below). We stress that in the sequel an incomplete probability vector will be used to represent the *same* state of knowledge as the corresponding open-frame body of evidence whose focal sets are all singletons (cf note 2).

We wish to draw the attention of the reader to an extreme case of (non-)knowledge, i.e. the case of incomplete bodies of evidence when there are *no focal sets* at all. It is obtained when one Dempster-pools two bodies of evidence which are *totally dissonant*, in the sense that any focal set in the first body has void intersection with any focal set in the second. We find it suggestive to dub this situation of a *totally self-dissonant* body of evidence as *total disconcertment*, to distinguish it from *total ignorance*, a closed-frame evidence when the only focal set is $\mathcal{X}$, $m(\mathcal{X}) = 1$.

3.4 The Recipe: Incomplete Interval Probabilities

If $\mathcal{P}$ is a closed and convex set of *incomplete* probability vectors, one can define the *incomplete lower probability* $P_*(A)$ as a minimum over $\mathcal{P}$ and the *incomplete upper probability* $P^*(A)$ as a maximum, exactly in the same way as in (2). One soon checks that P_* and P^* are both monotonic set-functions as in (1), P^* dominating P_*; P_* is superadditive, while P^* is subadditive. More than that, one soon proves that $A \cap B = \emptyset$ implies the four inequalities to follow:

$$P_*(A) + P_*(B) \leq P_*(A \cup B) \leq P_*(A) + P^*(B)$$
$$\leq P^*(A \cup B) \leq P^*(A) + P^*(B) \qquad (7)$$

$P_*(\mathcal{X})$ will be called the *lower load* of $\mathcal{P}$, while $P^*(\mathcal{X})$ will be its *upper load*. If all the probability vectors in $\mathcal{P}$ have the same load $P(\mathcal{X})$, we shall say that $\mathcal{P}$ is *flat*; then *duality* holds:

$$P_*(A) + P^*(\overline{A}) = P_*(\mathcal{X}) \equiv P^*(\mathcal{X}) \quad \forall A \qquad (8)$$

Observe that we are relaxing the usual definition of duality, where one requires that the sum be 1, rather than $P_*(\mathcal{X}) = P^*(\mathcal{X}) \leq 1$. Mathematically, flat families are nothing really new with respect to complete interval probabilities: just *normalize*, i.e. divide everything by the constant load $P(\mathcal{X})$, assumed positive (we stress that in general incomplete interval probabilities do *not* obey any duality rule, when they are not bound to be flat). More generally, flatness over C, i.e. $P_*(C) = P^*(C)$, holds iff all vectors P in $\mathcal{P}$ have constant load $P(C)$ over C.

One can prove that, at least formally, *open-frame evidence theory is a sub-case of incomplete interval probabilities* exactly in the same way as one proves that closed-frame evidence theory is a sub-case of complete interval probabilities [3]; the interval probability one arrives at is *always* flat with load $P(\mathcal{X}) = \sum_F m(F)$. More precisely, incomplete beliefs and plausibilities can be obtained as minima and maxima over the family $\mathcal{P}$ of incomplete probability vectors P such that:

$$\text{Bel}(A) \;\leq\; P(A) \;\leq\; \text{Pl}(A)$$

We find it useful to hint at the easy proof: if one "splits" an incomplete weight m over the singletons of F (this can be achieved in infinite ways, unless F is itself a singleton), and then, for each element a_i of $\mathcal{X}$, one sums up the splits inherited from the focal sets F where a_i belongs, one obtains an incomplete probability vector P with load equal to $\sum_F m(F)$. Now, $\mathcal{P}$ is equal to the family of incomplete probability vectors which are obtainable in this way; one soon checks that $P_* = \text{Bel}$, $P^* = \text{Pl}$ (to achieve $\text{Bel}(A)$, give F-shares to elements of A only when the focal set F is included into A; to achieve $\text{Pl}(A)$, give all F-shares to elements of A whenever the focal set F intersects A).

The (formal) inclusion between the two theories is strict, even if the interval probabilities are constrained to be flat, as easily shown [3]. Unlike in the general case of interval probabilities, beliefs are always *strongly* superadditive, and plausibilities are always *strongly* subadditive. As soon checked:

Proposition 1 *A body of evidence is flat over C iff there is no focal set F which intersects both C and $\overline{C}$.*

Incomplete probability vectors *à la Rényi*, or better their epistemic counterparts, can be identified with incomplete interval probabilities whenever $|\mathcal{P}| = 1$, and so the promised unification has been achieved: complete interval probabilities, complete and incomplete bodies of evidence and incomplete point probabilities are all special cases of incomplete interval probabilities, at least *formally*. The "elementary building blocks" which the new formalism imposes on us are incomplete probability vectors, which we have assumed to be legitimate descriptions for a state of knowledge. To those who see the inclusion *closed-frame evidence theory* $\subset$ *closed-frame interval probabilities* as purely formal, the corresponding open-frame inclusion will appear to be quite unappealing; however, the intersection is not void even at the conceptual level, as soon as one is ready to identify "Rényian" bodies of evidence (focal sets are singletons) with open-frame interval probabilities for which $|\mathcal{P}| = 1$. Two of the current interpretations of interval probability are so "general-purpose" that to our mind they extend also to our open-frame setting:

i) *unpooled opinions*: several experts have described their states of knowledge through incomplete probability vectors; we wish to keep their opinions *unpooled*, so as to know which is the minimum incomplete probability and which is the maximum incomplete probability given to each event by the team of experts as a whole (presumably the number of experts will be finite, in which case $\mathcal{P}$ will be the convex hull of a finite number of "points", or incomplete vectors)

ii) *approximation*: the "true" state of knowledge is an incomplete probability vector which we are unable to describe exactly; $\mathcal{P}$ simply contains all its possible descriptions (cf example 3 below)

The following examples are taken from [3]; example 2 is re-worked:

Example 1 Assume the agent (the expert, and at the same time the experimenter) knows *a priori* that there are K distinct observables and n experiments are made. Some of the experiments are however "hazy". If a_i is observed n_i times, the observed relative frequencies give rise to an incomplete probability vector $P = \{\ldots, \frac{n_i}{n}, \ldots\}$ with $n_T = \sum_i n_i < n$; the corresponding state of knowledge fits into closed-frame evidence theory, by setting $m(a_i) = \frac{n_i}{n}$, $m(\mathcal{X}) = 1 - \frac{n_T}{n}$, so that $\mathrm{Bel}(a_i) = \frac{n_i}{n}$, $\mathrm{Pl}(a_i) = \frac{n_i}{n} + m(\mathcal{X}) = 1 - \sum_{j \neq i} \frac{n_j}{n}$. The interpretation of Bel and Pl in terms of frequencies is obvious: had one been able to see through the haze, the observed frequencies would have fallen between $\frac{n_i}{n}$ and $1 - \sum_{j \neq i} \frac{n_j}{n}$, i.e. somewhere in between beliefs and plausibilities.

Example 2 One assigns the missing mass $1 - \sum_i \frac{n_i}{n}$ to an extra-element z which has been appended to the closed frame $\mathcal{X}$ of example 1; z is a sort of *tag* meant to take into account all the observables which are still unknown, i.e. the *unknown propositions* as in [6]. One obtains an incomplete, or self-dissonant, state of knowledge if one Dempster-pools with a complete unifocal body of evidence m_1 for which $m_1(\mathcal{X}) = 1$. In the language of [6], one is Dempster-conditioning over the frame $\mathcal{X}$ of all *possible* and *impossible* propositions, whose (mere) existence is well-known to the agent (in section 4 we shall describe *two* forms of Dempster conditioning, but they coincide in this case, as follows from theorem 1 below). So doing, we have curtailed plausibilities, which are now as small as the beliefs: $\mathrm{Pl}(a_i) = \mathrm{Bel}(a_i) = \frac{n_i}{n}$; with respect to example 1, we got rid of "ignorance" (intervals boiled down to points), at the price however of replacing it with incompleteness. The new situation corresponds to the open-frame body of evidence defined by setting $m(a_i) = \frac{n_i}{n}$, while $m(\mathcal{X}) = m(z) = 0$. This example is crucial to assess the purport of the theory; note there is no need to think that the incomplete vector is a "second-step" state of knowledge: the mechanism of Dempster pooling might have taken place, possibly unconsciously, inside the brain of one and the very same agent.

Example 3 Assume the data are communicated in a slightly imprecise way, up to an error $\epsilon > 0$: $f_i = \frac{n_i}{n} \mp \epsilon$. To take this into account one may replace point values by small intervals: $\mathcal{P} = \{P \,:\, f_i - \epsilon \leq p_i \leq f_i + \epsilon,\ \sum f_i - \delta \leq \sum_i p_i \leq \sum_i f_i + \delta\}$, the last constraint being useless when $\delta \geq K\epsilon$. Such an approximate state of knowledge is not flat, and as such lies outside the capacity of incomplete bodies of evidence (for a specification of the consistency constraints to impose on ϵ and δ, cf [3]). Again, "noisy communication" might have taken place *inside* the brain of the agent.

4 Slicing versus Open-frame Dempster Conditioning

As usual in evidence theory, we interpret conditioning a state of knowledge with respect to event C as pooling that state of knowledge with respect to the closed-frame evidence $m_2(C) = 1$. Even so, we are given *two* alternatives: either conditioning separately each "building-block" $P \in \mathcal{P}$, or conditiong the state of knowledge taken as a whole. The latter is one of the standard ways of conditioning, especially when the state of knowledge to be conditioned is a complete body of evidence, and when one uses the closed-frame Dempster rule with the normalization coefficient; the first is more directly inspired by our new approach.

We begin by the first form of open-frame Dempster conditioning, which, to avoid confusion with the second, we call *slicing*. Consider the new *sub-universe* (conditioning event) C, $\emptyset \neq C \subseteq \mathcal{X}$, and set:

$$\mathcal{P}_C = \{P_C : P_C(A) = P(A \cap C),\ P \in \mathcal{P}\}$$

The vectors in $\mathcal{P}_C$ are the same as those in $\mathcal{P}$, after turning to zero the components p_i relative to elementary events $a_i \notin C$. It is as if $\mathcal{X} - C$ were "lost into haze"; less imaginatively: to each "building block" $P \in \mathcal{P}$ we are separately appling Dempster rule with respect to the body of evidence $m_2(C) = 1$. Not to overcharge our notation, we shall *not* use any symbol for conditioning, and write simply $P_*(A \cap C)$ and $P^*(A \cap C)$. Notice that "sliced" lower and upper probabilities are obtainable as extrema on the new family $\mathcal{P}_C$, and so have all the formal properties which pertain to usual lower and upper incomplete probabilities. Unless $P_*(C) = 1$, slicing turns closed-frame knowledge into strictly open-frame knowledge.

Now we go to the second form of open-frame Dempster conditioning. If $\mathcal{P}$ is derived from (formally equal to) a body of evidence m, one can apply directly Dempster rule with respect to $m_2(C) = 1$, and express the resulting beliefs and plausibilities as in (6) (we stress that these expressions can be obtained directly, *without using duality*). By a formal analogy, we can extend the expressions of (6) to any incomplete interval probability $\mathcal{P}$ to obtain a new kind of *open-frame Dempster conditioning* for any incomplete interval probability, to be compared with slicing. We shall set for any family $\mathcal{P}$, but only formally when $\mathcal{P}$ is *not* a body of evidence:

$$P_*(A|C) = P_*(A \cup \overline{C}) - P_*(\overline{C})\ ,$$

$$P^*(A|C) = P^*(A \cap C) \qquad (9)$$

The upper conditional probability is the same as for slicing, and so can be obtained as a maximum over $\mathcal{P}_C$. Unfortunately (but not surprisingly), the lower conditional probability does *not* coincide with $P_*(A \cap C)$, as given by slicing. One has only:

$$P_*(A \cap C)\ \leq\ P_*(A|C)$$

as soon follows from the superadditivity of P_* applied to $A \cap C$ and $\overline{C}$.

Below, a comparison between slicing and open-frame conditioning is carried on, so as to understand when the two are one and the same thing. We first list some basic properties of $P_*(A|C)$; up to the last, they do not assume flatness of $\mathcal{P}$, the interval probability to be conditioned. It turns out that $P_*(\ |C)$ gives a confidence distribution over the sub-frame C (and also over the entire frame $\mathcal{X}$), which is dominated by $P^*(\ |C)$:

Properties

$$P_*(\emptyset|C) = 0,\ \ P_*(A|C) = P_*(A \cap C|C)$$

$$P_*(C|C) = P_*(\mathcal{X}|C) \leq 1,\ \ P_*(A|\mathcal{X}) = P_*(A)$$

$$A \subseteq B\ \Rightarrow\ P_*(A|C) \leq P_*(B|C)$$

$$P_*(A|C) \leq P^*(A|C)$$

$$P_*(C)\ \leq\ P_*(C|C)\ \leq\ P^*(C)$$

$$\textit{if } \mathcal{P} \textit{ is flat}\ \ P_*(C|C) = P^*(C)$$

Proofs soon follow from definitions (9); to prove $P_*(A|C) \leq P^*(A|C)$ use the second inequality in (7) applied again to $A \cap C$ and $\overline{C}$. If $\mathcal{P}$ is flat the inequality $P_*(C|C) \leq P^*(C)$ can be strengthened to equality by using the definition of $P_*(C|C)$ and duality (8). The following corollary is soon obtained:

Corollary 1 *If the family $\mathcal{P}$ is flat, any conditional $P_*(\ |C)$ is flat, whatever the conditioning event C; if $\mathcal{P}$ is flat over C, the conditional $P_*(\ |C)$ is flat.*

Proof.

The first implication is a re-phrasing of the last property. As for the second implication, use the property before the last. In both cases recall that $P^*(C) = P^*(C|C)$.

Now we come to a property which unfortunately does not hold in general, that is superadditivity of $P_*(\ |C)$; superadditivity does hold, however, when the unconditional P_* is *strongly* superadditive, e.g. because it is derived from a body of evidence m and so can be interpreted as a belief:

Proposition 2 *The conditional $P_*(\ |C)$ is superadditive whatever the conditioning event C iff the unconditional P_* is strongly superadditive.*

Proof. Without restriction assume $A \cup B \subseteq C$. One has to prove: $P_*(A \cup \overline{C}) + P_*(B \cup \overline{C}) \leq P_*(A \cup B \cup \overline{C}) + P_*\big([A \cap B] \cup \overline{C}\big)$; this is precisely strong superadditivity on $A \cup \overline{C}$ and $B \cup \overline{C}$.

Theorem 1 *If the family $\mathcal{P}$ is flat, the following statements are equivalent:*

i) open-frame Dempster conditioning over C coincides with slicing over C

ii) $\mathcal{P}$ is flat over C

Proof.

i) $\Rightarrow$ ii) Take $A = C$ to obtain $P_*(C|C) = P_*(C)$, i.e. $P_*(\mathcal{X}) = P_*(C) + P_*(\overline{C})$; this, by duality (8) applied to $P_*(\overline{C})$, becomes $P_*(C) = P^*(C)$.

ii) $\Rightarrow$ i) First observe that flatness over $\mathcal{X}$ and C implies flatness over $\overline{C}$. Now, flatness over $\overline{C}$ always implies i) (use the first two inequalities in (7) with $A \cap C$ and $\overline{C}$ rather than A and B).

If $\mathcal{P}$ is not flat, neither does i) imply ii) (just take $C = \mathcal{X}$), nor does ii) imply i), as shown by the following example:

Example Take $\mathcal{X} = \{a, b, c\}$, $C = \{a, b\}$, $A = \{a\}$, $\mathcal{P}$ equal to the segment with endpoints $P = (\frac{1}{3}, \frac{1}{4}, \frac{1}{4})$ and $Q = (\frac{1}{4}, \frac{1}{3}, \frac{1}{3})$. One has $P_*(A|C) = \frac{1}{3} > P_*(A \cap C) = \frac{1}{4}$.

Proposition 1 and theorem 1 soon imply:

Corollary 2 *In the case of a body of evidence, open-frame Dempster conditioning over C coincides with slicing over C iff there is no focal set intersecting both the conditioning event C and its negation $\overline{C}$.*

5 Final Comments

Incomplete interval probabilities, as expounded in [3], originated as a purely formal proposal. Actually, one of the arguments, if not the most solid one, which is used to maintain that evidence theory is not a subtheory of interval probabilities, is precisely that the latter have to stop when evidence theory can instead go on, by "puncturing" the frame so as to let in what Smets calles "unknown propositions", to be added to the possible and the impossible ones [6]: we have shown that this formal objection falls, as soon as one is ready to make use of a nice mathematical object called an incomplete probability vector (or an incomplete probability distribution), which has a respectable pedigree, since it goes back to no less but Alfred Rényi. This as far as form is concerned; should it turn out that our proposal is only a formal one, all the same it would help to understand the relation, or the lack of relation, between probability and evidence. As a historical remark, Rényi himself did come across incomplete probability vectors because of merely formal snags, but soon got interested in "filling them up" with proper contents [1], and so became a forerunner of "unorthodox probabilities", even if his point of view was not an epistemic one.

Let us turn to contents. The basic demand is this: is there any place for incompleteness or contradictoriness also inside interval probabilities, which have already proved able to represent and manage not only uncertainty, as does the usual pointwise Bayesian theory, but also ignorance and imprecision, as happens in [complete] interval probabilities (and also in closed-frame evidence theory)? Incompleteness and contradictoriness are indispensable facets of human reasoning, which have been brilliantly brought into evidence theory by Smets, when he showed how to open the frame of reference, and, by so doing, cut the link between evidence and probabilities, at least apparently (we refer to [6] for a discussion).

We come to our proposal. In a way, one might even object that our "new" notions are not so new, after all: they just re-formalize and somewhat extend old notions, and so do not really need any new defence. To begin with, Rényian incomplete probabilities *already* have a place in open-frame evidence theory, when the focal sets are singletons: actually, it is not a great requirement to agree that such a body of evidence and the corresponding Rényian probability vector represent the very same state of knowledge, this being little more than a convention (compare with note 2). On the other hand, even if (epistemic-flavoured) Rényi's incomplete probability vectors are given a right of citizenship by this unassuming convention, this is not enough to accept them as the "elementary components" of a theory such as ours; the stumbling block is precisely this. In both open-frame and closed-frame evidence theory the "elementary building blocks" are felt to be unifocal bodies of evidence, while in complete interval probabilities this role is taken on by Bayesian probability vectors; the new building blocks generalize the latter. We feel that a philosophical justification of our proposal goes necessarily through a careful evaluation of the epistemic meaning of a Rényian probability distribution, as meant to take into account uncertainty and contradictoriness (self-dissonance) *when there is no ignorance* (no imprecison). Rényian incomplete probability vectors do represent special states of knowledge, namely those when confidence intervals boil down to points: this is the feature which "Rényian probabilities", i.e. the epistemic counterparts to the incomplete frequentist probabilities put forward by Rényi, have in common with Bayesian probabilities. Now, the decomposition of incomplete bodies of evidence into Rényian probability vectors such that

$$\mathrm{Bel}(A) \ \leq P(A) \ \leq \mathrm{Pl}(A) \quad \forall A$$

and the fact that each incomplete belief is a minimum over such vectors, while each plausibility is a maximum, are mathematical facts, very easy to prove, on top of that (cf subsection 3.4). If the body of evidence is complete, the Rényian probabilities in its decomposition happen to be also Bayesian probabilities; some see this as a meaningful decomposition into pointwise, not interval-like, states of knowledge. Acceptance of this point of view is a preliminary step to the acceptance of the general decomposition into Rényian probabilities as a meaningful one, and the general decompositions unavoidably point to an extension when the family $\mathcal{P}$ of probability vectors is not necessarily derived from an incomplete body of evidence m.

The two ways of conditioning considered in this paper, one being inspired by usual arguments about bodies of evidence, while the other, i.e. slicing, appears to be more internal to probabilities and nearer to Bayes conditioning, may help understanding the purport of incomplete probabilities. As far as decisions are concerned, i.e. when one goes to the *pignistic* level, presumably incomplete states of knowledge have to be first "forced" into completeness along lines as those shown by Smets for bodies of evidence [5]; cf also [3]. This, however, is a step to be taken at a later stage, when and if our approach will be better vindicated.

References

[1] J. Aczél. Personal communication

[2] A. Rényi. On Measures of Entropy and Information. In *Selected Papers of Alfréd Rényi*, 2:565-579. Akadémiai Kiadó, Budapest, 1976.

[3] A. Sgarro. An Open-frame Theory of Incomplete Interval Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6.6, 1998

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

[5] Ph. Smets. Constructing the Pignistic Probability Function in a Context of Uncertainty. *Proc. Fifth Workshop Uncertainty in AI*, Windsor, Canada, 319-326, 1989.

[6] Ph. Smets. The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12.5:447-458, 1990.

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

[8] Zh. Wang, G.J. Klir. *Fuzzy Measure Theory.* Plenum Press, 1992.

[9] L. Zadeh. A Mathematical Theory of Evidence (Book Review). *AI Magazine*, 5.3:81-83, 1984