Incremental Conditioning of Lower and Upper Probabilities

Lonnie Chrisman

Carnegie Mellon University
Pittsburgh, Pennsylvania

ABSTRACT

Bayesian-style conditioning of an exact probability distribution can be done incrementally by updating the current distribution each time a new item of evidence is obtained. Many have suggested the use of lower and upper probabilities for representing bounds on probability distributions, which naturally suggests an analogous procedure of incremental conditioning using forms of interval arithemetic. Unfortunately, conditioning of lower and upper probability bounds looses information, yielding incorrect bounds when updates are performed incrementally and making the conditioning operation non-commutative. Furthermore, when lower probability functions are represented by way of their Möbius transforms, the operation of conditioning can cause an exponential explosion in the number of non-zero Möbius assignments used to represent the function. This paper presents an alternative representation for lower probability that overcomes these problems. By representing the results of both Dempster conditioning and strong conditioning, the representation indirectly encodes lower probability bounds in a form that allows updates to be performed incrementally without a loss of information. Conditioning with the new representation does not depend on the order of updates or on whether evidence is incorporated incrementally or all at once. The bounds obtained are exact when the original lower probabilities satisfy a property called 2-monotonicity. Although the new representation encodes more information about probability bounds than the straight representation, updates on the new representation never increase the number of Möbius assignments used to encode the

Address correspondence to Lonnie Chrisman, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pa 15213-3891; e-mail: chrisman@cs.cmu.edu

International Journal of Approximate Reasoning 1995 13:1–25

655 Avenue of the Americas, New York, NY 10010

lower probability — a considerable improvement over the worst-case exponential increase seen with the straight representation. The new representation helps to improve the efficiency and convenience of representing and manipulating lower probabilities.

1. Introduction

The Bayesian probabilistic framework provides a methodology for reasoning about uncertainty ([31], [16], [21], [36]). Belief is represented by a single probability distribution and conditioning serves as the primary tool for updating belief as new information is obtained. An important characteristic of the updating process is that it can be done incrementally. In other words, each time a new fact is learned about the current situation, the probability distribution representing belief can be replaced by an updated distribution without any loss of information about the true situation[1].

The use of a single exact probability distribution in the pure Bayesian framework is often challenged. Some feel that an exact distribution fails to satisfactorily distinguish between uncertainty and ignorance or between certainty and confidence ([51], [26], [28], [27], [30], [61], [56], [39], [11], [40], [55], [53]). Others point out that often insufficient knowledge is available or that it is too time-consuming to obtain the necessary knowledge to warrant the precision inherent in exact probabilities ([17], [52], [54], [11], [2], [33], [29], [14]). In response to these objections and others, many researchers have suggested replacing the use of an exact probability distribution with probability intervals, where the intervals are specified by lower and upper bounds. Lower and upper probability bounds have also been found useful in the context of a traditional Bayesian framework for approximate computation ([8]), and to achieve a greater level of robustness ([60], [22, Chapter 10]).

Lower probabilities can be used to represent probability bounds ([41], [18], [12], [59], [5], [38], [14]). Let Ω denote a set of mutually exclusive and exhaustive situations. A lower probability, $\underline{P}$, is a function $\underline{P} : 2^{\Omega} \rightarrow [0,1]$ (satisfying certain conditions to be discussed in Section 2) that represents lower bound constraints on probability distributions. A probability distribution, P, is *consistent* with $\underline{P}$ if for every $A \subseteq \Omega$, $\underline{P}(A) \leq P(A)$. $\underline{P}$ can therefore be viewed as representing a set of probability distributions — namely, those that are consistent with it. Probability distributions, belief functions[2] ([41], [26], [15]), lower envelopes ([57], [34]), inner and outer

[1] See, for example, Proposition 3.1 in [20].

[2] While Belief Functions are mathematically instances of Lower Probability functions, they are often used in a manner inconsistent with a Lower Probabilistic interpretation.

measures ([15]) and probability bounds on individual elements of Ω ([17], [56]) are all special cases of lower probabilities.

Using a lower probability representation, it is natural to attempt inference in an analogous manner to Bayesian inference. Beginning with a lower probability function, one would incrementally update this function as new evidence arrives. But unlike the case with exact probabilities, a lower probability representation alone is not sufficient for capturing all the information about the current situation that is available at the time of update. In other words, each lower probability update looses information, such that bounds obtained after two or more successive updates may disagree with the bounds that would be obtained by conditioning the original belief in a single step with all available evidence ([37]). In fact, in general the bounds obtained will depend upon the order in which updates are performed ([20]). This loss of information is not a result of any particular conditioning rule, but is a result of the fact that the representation of lower probability is not powerful enough to capture all the information that is available at the time of an update ([24]).

In addition to the loss of information problem, a second problem can impede the use of lower probability intervals. In practical applications, one is often interested in "sparse" lower probability functions — representations where probability bounds can be described with a small number of mass-assignments. This is because in the general (non-sparse) case, the number of parameters in a lower probability specification can be enormous. For example, if $N = |\Omega|$ is finite, as many as 2^N numbers may be required. To take advantage of sparsity, one can represent the Möbius Transform of a lower probability function inside a computer rather than explicitly storing the function itself ([42]). In Möbius space, each non-zero set assignment can be viewed as one constraint, and a lower probability function is sparse when the overwhelming majority of Möbius set assignments are zero. A problem with the use of standard lower probability representations is that the conditioning of these representations does not preserve sparsity. After an update, the number of mass-assignments necessary to represent the function can grow considerably — increasing exponentially in some cases.

This paper introduces an alternative representation and method for conditioning lower probability functions that addresses the above two difficulties. It allows for incremental updating of lower probabilities without a loss of information, and when a sparse function is updated, the result remains sparse. Furthermore, it is convenient to perform the update in Möbius space.

Much of the existing literature concerning lower and upper probabil-

---

Both [41] and [20] discuss how belief functions may have either an evidential interpretation, as in the Dempster-Shafer theory of evidence ([40], [42]) and Transferable Belief Model ([50], [49]), or a lower probabilistic interpretation.

ity centers around the Dempster-Shafer theory ([40]). Many have attempted to relate the theory to probability theory and/or lower probability interpretations[3] ([19], [26], [28]). In this context, an interesting aspect of the representation introduced here is a relationship that is highlighted between Dempster's Rule of Conditioning and lower probabilistic conditioning (or convex conditioning). In particular, it is found that Dempster's Rule of Conditioning indirectly captures part (more than half, but not all) of the information necessary to maintain lower and upper probability intervals.

Section 2 reviews background, terminology and known results concerning lower probability. Section 3 demonstrates with an example the loss of information that results when a straight lower probability representation is updated incrementally. A solution is presented in Section 4 where a new representation for lower probability information is introduced, and its representation in terms of the Möbius transform is given in Section 5 followed by an example in Section 6 to demonstrate how the new representation is used. In Section 7 the complexity of updates is examined, where it is shown that the straight lower probability representation does not take advantage of sparsity while the new representation does. We conclude in Section 8.

2. Lower Probability

We begin in this section by reviewing terminology and known results concerning lower probability. Many similar properties and terminology have been developed and utilized by [6], [41], [22, Chap. 10], [58], [59], [35], [5], [34], [24] and others.

Every probabilistic argument begins with a mutually exclusive and exhaustive set of possible situations, denoted by Ω and termed a *frame of discernment*. A probability distribution on Ω is an additive set-function $P : 2^{\Omega} \rightarrow [0,1]$ with $P(\emptyset) = 0$ and $P(\Omega) = 1$. It is additive in that for any $A, B \subseteq \Omega$ with $A \cap B = \emptyset$

$$P(A \cup B) = P(A) + P(B)$$

We denote the set of all probability distributions on Ω by $\mathcal{M}$.

Lower and *Upper Probability* functions, $\underline{P}$ and $\overline{P}$, are also set-functions on Ω satisfying the following properties for any $A, B \subseteq \Omega$ with $A \cap B = \emptyset$:

1. $\underline{P}\,(\emptyset) = \overline{P}\,(\emptyset) = 0$
2. $\underline{P}\,(\Omega) = \overline{P}\,(\Omega) = 1$

[3] Some work has also strived to remove any relation whatsoever to a probabilistic interpretation (e.g., [50], [48]).

3. $\underline{P}(A) + \overline{P}(\bar{A}) = 1$
4. $\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B)$ (Super-Additivity)
5. $\overline{P}(A) + \overline{P}(B) \geq \overline{P}(A \cup B)$ (Sub-Additivity)

where $\bar{A}$ denotes $\Omega - A$, the *complement* of A.

Property 3 above specifies that $\underline{P}$ and $\overline{P}$ are mutual conjugates, and therefore it is only necessary to store one or the other since either can be readily obtained from the other. It is always the case that $\underline{P}(A) \leq \overline{P}(A)$.

We say a probability distribution P is *consistent* with a lower probability $\underline{P}$ (and implicitly its conjugate $\overline{P}$) when for every $A \subseteq \Omega$, $\underline{P}(A) \leq P(A)$. We denote by $\mathcal{P}(\underline{P})$ the set of all distributions consistent with $\underline{P}$. The above conditions that define lower probabilities are not strong enough to ensure that $\mathcal{P}(\underline{P}) \neq \emptyset$.

Suppose there exists a non-empty $\mathcal{P} \subseteq \mathcal{M}$ such that for all $A \subseteq \Omega$

$$\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A)$$
$$\overline{P}(A) = \sup_{P \in \mathcal{P}} P(A)$$

Then $\underline{P}$ is called a *lower envelope* and $\overline{P}$ is its conjugate, called an *upper envelope*. Every lower envelope is also a lower probability, but the converse does not hold.

Every lower probability function is *monotone* (sometimes called $1-$ *monotone*), meaning that

$$A \subseteq B \Rightarrow \underline{P}(A) \leq \underline{P}(B)$$

A stronger property called 2-*monotoncity* is often useful [6]. A lower probability $\underline{P}$ on Ω is $2-$ *monotone* when for every $A, B \subseteq \Omega$,

$$\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cap B) + \underline{P}(A \cup B)$$

Two-monotonicity is a sufficient (but not necessary) condition to ensure that $\underline{P}$ is a lower envelope. Also, it can be shown ([59]) that $\underline{P}$ is 2-monotone if and only if for every $A, B \subseteq \Omega$, $A \cap B = \emptyset$, there exists $P \in \mathcal{P}(\underline{P})$ such that

$$P(A) = \underline{P}(A) \text{ and } P(B) = \overline{P}(B) \quad (1)$$

The extra property of 2-monotonicity is often quite useful, particularly because it is often the weakest condition necessary for obtaining simple but exact closed form formulas for various inferences. It is the strongest property we will utilize for the results in this paper.

For the remainder of this paper, we will only consider finite frames of discernment (i.e., $|\Omega| < \infty$). Suppose Ω is a finite frame of discernment, and $\underline{P}$ is a lower probability function defined on Ω. The *Möbius Transform* of $\underline{P}$ is the set function $m : 2^{\Omega} \rightarrow \mathcal{R}$ defined by ([40, pg. 39])

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \underline{P}(B) \tag{2}$$

If $m(A) \geq 0$ for all $A \subseteq \Omega$, then $\underline{P}$ is called a *Belief Function*, and readers familiar with Dempster-Shafer theory will recognize m as the mass-assignment function; however, it is not required for $\underline{P}$ to be a belief function for the Möbius Transform to be defined, and in general, $m(\cdot)$ may be negative on some sets. Belief Functions are equivalent to what [6] termed ∞-monotone capacities ([55]). Every belief function is also 2-monotone, and therefore is a lower envelope, and therefore is a lower probability, but 2-monotone lower envelopes are not, in general, belief functions.

The sets with non-zero Möbius assignments are termed the *focal elements* of $\underline{P}$.

The Möbius Transform is information preserving, such that the original function $\underline{P}$ can be recovered from m using the Inverse Möbius Transform given by

$$\underline{P}(A) = \sum_{B \subseteq A} m(B) \tag{3}$$

The original proof of this inverse relationship was given by [40, Theorem 2.2] in the context of belief functions, but his proof did not rely on the non-negativity of $m(\cdot)$ so it holds for more general lower probability functions as well. The same proof is rewritten in terms of the more general case in [5, Appendix] (see also [25]). The upper probability function is also readily available from the Möbius Transform using

$$\overline{P}(A) = 1 - \underline{P}(\bar{A}) = \sum_{B \not\subseteq \bar{A}} m(B) \tag{4}$$

We can interpret each non-zero Möbius assignment as a constraint on the allowable probability distributions. A positive assignment, $m(A) = x$, specifies that x units of probability mass is constrained to the set A, but within A may be redistributed arbitrarily. A negative mass assignment, $m(A) = -x$, specifies that x units of probability anti-mass is constrained to the set A and can be redistributed onto the positive probability mass within A so as to cancel out an equivalent amount of positive probability mass. We can therefore view the number of focal elements as a measure of the number of constraints specifying the bounds in a lower probability function. Other forms of constraints are also possible, but are not considered in this paper.

The most general lower probability requires 2^N numbers ($N = |\Omega|$) to specify either $\underline{P}$ or its Möbius Transform m. For any sizable domain, this is prohibitive; however, for many applications where lower probabilities might be of interest, the probability bounds arise from a relatively small number of mass assignments. It is therefore typically most convenient to represent m inside a computer, rather than $\underline{P}$, since only the non-zero assignments must be enumerated. When the number of focal elements is small, we say that $\underline{P}$ is *sparse*.

2.1. An Example

We introduce a simple example here which will be used to demonstrate some of the basic ideas. The example is entirely hypothetical and is not intentionally based on any accepted paleontological fact.

A group of paleontologists are beginning to hunt fossils in a previously unexplored but very unusual region of northwest Burkawaland (a ficticious place). The only fossils they expect to find are those of mammals, birds, reptiles and fish, but because they know so little about the area and because the region is so unusual, they are uncomfortable with the idea of estimating an exact *a priori* distribution over fossil types. They choose instead to estimate a prior lower probability function.

The only previous study of the area stated that out of 100 fossils that had been examined, three were determined to be of mammalian origin, and 50 were clearly from fish. There were 46 specimens that were believed to be either a species of fish or reptile, but with the tools available at the time of the study, there was no way to determine which. Finally, there was one specimen that might have been either a bird or a reptile, but again, that could not be determined. Based solely on this study, by accepting the proportions found in the study as being indicative of the population of fossils as a whole, the group adopts the lower probability distribution whose Möbius Transform is the following:

$$\begin{aligned} \Omega &= \{mammal, bird, reptile, fish\} \\ m_0(\{mammal\}) &= 0.03 \\ m_0(\{fish\}) &= 0.5 \\ m_0(\{reptile, fish\}) &= 0.46 \\ m_0(\{bird, reptile\}) &= 0.01 \\ m_0(\text{all other sets}) &= 0 \end{aligned} \tag{5}$$

This function specifies probability bounds — for example $P_0(\{bird, fish\})$ is bounded by $\underline{P_0}(\{bird, fish\}) = 0.5$, and $\overline{P_0}(\{bird, fish\}) = 0.97$. The example happens to be a belief function (and therefore, also a 2-monotone lower envelope).

3. Incremental Conditioning

Conditioning is the primary mechanism for incorporating evidence within a Bayesian framework. In this framework, one begins with knowledge about uncertainty explicitly encoded in the form of an *a priori* probability distribution P_0. After learning that $E_1 \subseteq \Omega$ is true, an updated belief, $P_1 : 2^{\Omega} \rightarrow [0, 1]$, is obtained using

$$P_1(A) = P_0(A|E_1) = \frac{P_0(A \cap E_1)}{P_0(E_1)}$$

This is the definition of conditional probability. When a second item of evidence is obtained, E_2, the process repeats using the first updated belief as the starting point:

$$P_2(A) = P_1(A|E_2) = \frac{P_1(A \cap E_2)}{P_1(E_2)} \qquad [= P_0(A|E_1, E_2)]$$

Conditional probability has the very important property that the same result is obtained independent of the order of updates and whether or not updates are performed incrementally.

Consider the same incremental process stating with the lower probability function in (5) from the example of Section 2.1. Suppose a new fossil is discovered, and the team now wishes to determine its type (and associated uncertainty). First, a team member notes that the animal had legs, and therefore was not a fish ($E_1 = \{mammal, bird, reptile\}$). What does it mean to update the lower probability function? Ideally, the result should represent the envelope obtained by collecting each probability distribution consistent with the original function after it has been updated with the new evidence. This desired envelope after learning $E_1 = \{mammal, bird, reptile\}$ is given by the following (again, the Möbius Transform is shown):

$$\begin{aligned} m_1(\{mammal\}) &= 0.06 \\ m_1(\{mammal, reptile\}) &= 0.69 \\ m_1(\{bird, reptile\}) &= 0.25 \end{aligned} \tag{6}$$

Later it will be shown how such computations can be performed. The bounds given by this function are tight — for any set $A \subseteq \Omega$, one can find a probability distribution consistent with the initial belief that yields the bound after conditioning. For example, the bound $\underline{P}(\{mammal, reptile\}) = 0.75$ is obtained from the distribution $\langle .03, .01, 0, .96 \rangle$.

Suppose another test reveals that the specimen was definately not that of a reptile, thus indicating that the classification is in $E_2 = \{mammal, bird, fish\}$. If (6) is updated, as one would do using incremental updating, the following Möbius function is obtained:

$$\begin{aligned} m_2(\{mammal\}) &= 0.19 \\ m_2(\{mammal, bird\}) &= 0.81 \end{aligned} \tag{7}$$

One would hope that by updating the original belief, m_0, in one step with all the evidence learned thus far, $E = E_1 \cap E_2 = \{mammal, bird\}$, the same result would be obtained. Unfortunately, this is not the case. Updating m_0 with $\{mammal, bird\}$ yields:

$$\begin{aligned} m_3(\{mammal\}) &= 0.75 \\ m_3(\{mammal, bird\}) &= 0.25 \end{aligned} \tag{8}$$

Comparing (7) and (8) shows that they are actually quite a bit different (e.g., $\underline{P_3}\ (\{mammal\}) - \underline{P_2}\ (\{mammal\}) = 0.56$). One might also consider what happens if we perform the updates in the opposite order, first conditioning on $E_2 = \{mammal, bird, fish\}$ then on $E_1 = \{mammal, bird, reptile\}$. Doing so in this example yields yet another result:

$$\begin{aligned} m_4(\{mammal\}) &= 0.62 \\ m_4(\{mammal, bird\}) &= 0.38 \end{aligned}$$

The fact that all of these are different is very dissatisfying and is something that does not occur with exact probabilities. The problem was previously noted by [20] and [37], and similar problems for other forms of interval probabilities (e.g., Dempster-Shafer) have also been discussed in the literature ([1]).

The reason that incremental updating does not work is that the lower probability representation is not sufficiently powerful to represent all the information that is available ([24]). Each time an update is performed, some information is lost, and the information that is lost can be important for determining bounds for subsequent updates. In other words, some new consistent distributions are introduced which are not the result of the original set of consistent distributions. An analogy is useful for seeing how this is possible. Figure 1(a) shows a set of points in the plane which is exactly represented by the indicated bounds. After a transformation is applied to these points, the resulting bounds can again be perfectly represented, as shown in Figure 1(b), even though the bounds do not capture the set of points exactly. After a second transformation, Figure 1(c), the loss of information is reflected in the incrementally updated bounds.

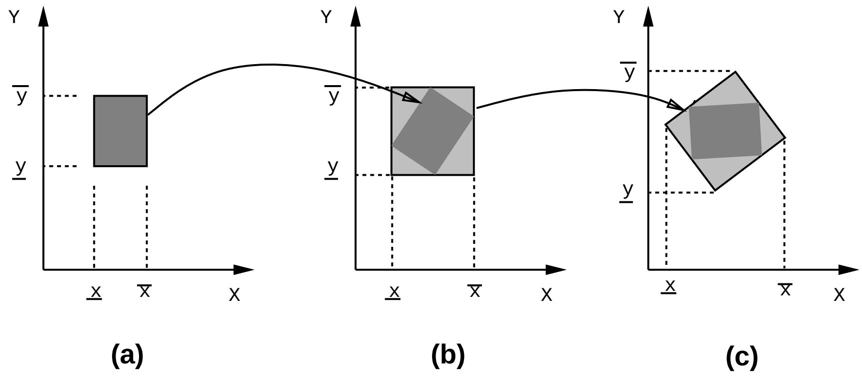

Figure 1. The loss of information from incremental updating of bounds.

Overcoming the loss of information with incremental updates requires a more powerful representation. One such representation that has been studied by [2], [56] and others is a system of linear constraints in the $(N-1)$-dimensional simplex of probability vectors, where the linear constraints specify a convex polytope with a finite number of sides. While such a representation does provide sufficient information to enable incremental updates, it is not as popular as lower probability representations, primarily because it is much more tedious to use and because the number of parameters in such a representation can quickly become unmanageable, as discussed in [56]. Walley [58] has also introduced a more general representation called lower previsions, which is equivalent to closed convex sets of probabilities ([22, Page 256], [58, Section 3.3]).

4. A New Representation

This section presents a new representation that is quite convenient to use, and which captures all the information necessary for obtaining correct bounds after incremental updating. When the starting lower probability function is 2-monotone, the bounds are guaranteed to be tight (i.e., the bounds will be achieved for some initial consistent prior probability distribution). The following section examines its Möbius Transform.

Suppose $\underline{P}$ is an initial lower probability function. After learning $E = E_1 \cap E_2 \cap \ldots$, the desired updated function (whether it is computed incrementally or not) is

$$\underline{P}(A|E) = \inf\{P(A|E) : P \in \mathcal{P}(\underline{P}), P(E) > 0\} \tag{9}$$

This form of conditioning has been referred to as *convex conditioning* ([26]) and the *cautious Bayesian* ([14]) approach to conditioning. Of course, it would not be feasible to enumerate all consistent probability distributions and then perform the above inf operation, but it is easy to obtain a bound for $\underline{P}(A|E)$. Recall the following rule of probability, where P is a probability distribution:

$$P(A|E) = \frac{P(A \cap E)}{P(A \cap E) + P(\bar{A} \cap E)} \tag{10}$$

Since for every set $B \subseteq \Omega$, $\underline{P}\,(B) \leq P(B) \leq \overline{P}\,(B)$ are lower and upper bounds on P, we can plug these bounds into (10) to obtain

$$P(A|E) \geq \frac{\underline{P}\,(A \cap E)}{\underline{P}\,(A \cap E) + \overline{P}\,(\bar{A} \cap E)} \tag{11}$$

Therefore, the right hand side of (11) is a lower bound. The lower bound is undefined when $\overline{P}\,(E) = 0$ and requires a somewhat different treatment, considered later, when $\overline{P}\,(E) > \underline{P}\,(E) = 0$. Recall that when $\underline{P}$ is 2-monotone, there exists a distribution P such that $P(A \cap E) = \underline{P}\,(A \cap E)$ and $P(\bar{A} \cap E) = \overline{P}\,(\bar{A} \cap E)$, so that the bound is tight when $\underline{P}$ is 2-monotone. This formula has been identified previously for the special case of belief functions by [10, Equation 4.8], [58, pg. 301], [38], [15], [60], [24], [49] and [9]. Additional properties for the belief function case are given in [15] and [24]. As discussed in the previous section, even when the bounds are tight, information may be lost.

Other forms of conditioning are possible, although each returns something different, so the results must be interpreted with caution. For example, Dempster [10] introduced the following conditioning rule:

$$\overline{P}\,(A||^{*}E) = \frac{\overline{P}\,(A \cap E)}{\overline{P}\,(E)} \tag{12}$$

It is undefined when $\overline{P}\,(E) = 0$. The rule is called *Dempster's Rule of Conditioning*. The notation $||^{*}$ is used here to distinguish this rule of conditioning from (11) and from other possible conditioning rules. The interpretation of this rule has been considered in [40], [44], [42], and [50] and is often intuitively viewed as a measure of evidential support — the degree to which the evidence supports the hypotheses. However, we can see that Dempster's Rule is not appropriate for directly describing lower probabilities as we desire since, for example, the bounds produced by Dempster's Rule are too narrow for our desired probabilistic interpretation ([10]):

$$\underline{P}\,(A|E) \leq \underline{P}\,(A||^{*}E) \leq P(A|E) \leq \overline{P}\,(A||^{*}E) \leq \overline{P}\,(A|E) \tag{13}$$

A dual to Dempster's conditioning rule is possible:

$$\underline{P}\,(A||_*E) = \frac{\underline{P}\,(A \cap E)}{\underline{P}\,(E)} \tag{14}$$

It is undefined when $\underline{P}\,(E) = 0$. This has been called *strong conditioning* ([32], [13]) and *geometric conditioning* ([55], [23], [47], [46]) and might be viewed as a measure of *evidential predictability* — the degree to which the initial belief predicts the evidence. An interpretation of this rule is considered in [47]. Like Dempster's rule, the bounds produced by this rule are too tight for the desired probabilistic interpretation.

Neither (12) nor (14) produce the lower envelope bounds that we are interested in. However, by plugging (12) and (14) into (11), we obtain

$$\underline{P}\,(A|E) \geq \frac{\underline{P}\,(A||_*E)\underline{P}\,(E)}{\underline{P}\,(A||_*E)\underline{P}\,(E) + \overline{P}\,(A||^*E)\overline{P}\,(E)} \tag{15}$$

This is defined when $\underline{P}\,(E) > 0$, and the bound is guaranteed to be tight when the original lower probability function is 2-monotone. In what follows, we will also handle the important case where $\overline{P}\,(E) > \underline{P}\,(E) = 0$. This rule is the basis for our new approach to conditioning.

Although the lower probability bounds given by $\underline{P}\,(A|E)$ and $\overline{P}\,(A|E)$ do not possess the representational power to prevent a loss of information, the information contained jointly within $\underline{P}\,(A||_*E)$ and $\overline{P}\,(A||^*E)$, along with two scalar values, do possess adequate information. These are not the actual bounds — but they indirectly encode all the information that is necessary to compute the desired bounds, using (15), for any proposition of interest. After observing E, we maintain four items: $\underline{P}\,(A||_*E)$, $\underline{P}\,(A||^*E)$, $\underline{P}\,(E)$, and $\overline{P}\,(E)$, the last two of which are simply scalars.

A first observation to make is that both $\underline{P}\,(A||_*E)$ and $\overline{P}\,(A||^*E)$ can be computed incrementally, as the following theorem demonstrates.

THEOREM 1. *Let $\underline{P}$ and $\overline{P}$ be lower and upper probability functions. For all sets $A, E_1, E_2 \subseteq \Omega$*

$$\underline{P}\,(A||_*E_1 \cap E_2) = \frac{\underline{P}\,(A \cap E_2||_*E_1)}{\underline{P}\,(E_2||_*E_1)} \tag{16}$$

or $\underline{P}\,(\cdot||_*E_1 \cap E_2)$ is undefined if $\underline{P}\,(E_2||_*E_1) = 0$ or if $\underline{P}\,(\cdot||_*E_1)$ is undefined. Also

$$\overline{P}\,(A||^*E_1 \cap E_2) = \frac{\overline{P}\,(A \cap E_2||^*E_1)}{\overline{P}\,(E_2||^*E_1)} \tag{17}$$

or $\overline{P}\,(\cdot||^{*}E_1 \cap E_2)$ *is undefined if* $\overline{P}\,(E_2||^{*}E_1) = 0$ *or* $\overline{P}\,(\cdot||_{*}E_1)$ *is undefined.*

Proof Saying $\underline{P}\,(\cdot||_{*}E_1)$ is undefined is equivalent to saying $\underline{P}\,(E_1) = 0$, and when this is the case, then from monotonicity, $\underline{P}\,(E_1 \cap E_2) = 0$. If $\underline{P}\,(E_2||_{*}E_1) = 0$ then by (14), $\underline{P}\,(E_1 \cap E_2) = 0$, so in either of these cases, $\underline{P}\,(\cdot||_{*}E_1 \cap E_2)$ is undefined. Otherwise, from definition (14):

$$\begin{aligned}
\underline{P}\,(A||_{*}E_1 \cap E_2) &= \underline{P}\,(A \cap E_1 \cap E_2)/\underline{P}\,(E_1 \cap E_2) \\
&= \frac{\underline{P}\,(A \cap E_1 \cap E_2)}{\underline{P}\,(E_2)} \Big/ \frac{\underline{P}\,(E_1 \cap E_2)}{\underline{P}\,(E_2)} \\
&= \frac{\underline{P}\,(A \cap E_1||_{*}E_2)}{\underline{P}\,(E_1||_{*}E_2)}
\end{aligned}$$

The proof of (17) follows the same form. ∎

A second observation to make is that $\underline{P}\,(E)$ and $\overline{P}\,(E)$ can also be computed incrementally when we maintain the four items of information mentioned previously. This follows directly from (14) and (12) as follows:

$$\underline{P}\,(E_1 \cap E_2) = \underline{P}\,(E_2||_{*}E_1) \cdot \underline{P}\,(E_1) \tag{18}$$

$$\overline{P}\,(E_1 \cap E_2) = \overline{P}\,(E_2||^{*}E_1) \cdot \overline{P}\,(E_1) \tag{19}$$

The following theorem shows that the four items of information ($\underline{P}\,(\cdot||_{*}E)$, $\overline{P}\,(\cdot||^{*}E)$, $\underline{P}\,(E)$, and $\overline{P}\,(E)$), which we now know can be updated incrementally, are sufficient for determining the lower (and upper) probability bounds.

THEOREM 2. *Let* $\underline{P}$ *and* $\overline{P}$ *be conjugate lower and upper probabilities. Then for* $A, E \subseteq \Omega$, *if* $\underline{P}\,(E) > 0$

$$\underline{P}\,(A|E) \geq \frac{\underline{P}\,(A||_{*}E)\underline{P}\,(E)}{\underline{P}\,(A||_{*}E)\underline{P}\,(E) + \overline{P}\,(\bar{A}||^{*}E)\overline{P}\,(E)} \tag{20}$$

or is undefined if $\overline{P}\,(E) = 0$. *If* $\overline{P}\,(E) > \underline{P}\,(E) = 0$, *then for all* $A \subseteq \Omega$

$$\underline{P}\,(A|E) \geq \begin{cases} 1 & \text{if } \underline{P}\,(A||^{*}E) = 1 \\ 0 & \text{if } \underline{P}\,(A||^{*}E) < 1 \end{cases} \tag{21}$$

If $\underline{P}$ *is 2-Monotone, then the bounds are tight, so that*

$$\underline{P}\,(A|E) = \frac{\underline{P}\,(A||_{*}E)\underline{P}\,(E)}{\underline{P}\,(A||_{*}E)\underline{P}\,(E) + \overline{P}\,(\bar{A}||^{*}E)\overline{P}\,(E)} \tag{22}$$

if $\underline{P}(E) > 0$ *and*

$$\underline{P}(A|E) = \begin{cases} 1 & \text{if } \underline{P}(A||^*E) = 1 \\ 0 & \text{if } \underline{P}(A||^*E) < 1 \end{cases} \tag{23}$$

if $\overline{P}(E) > \underline{P}(E) = 0$.

Proof Consider $\underline{P}(E) > 0$. Let $P \in \mathcal{P}(\underline{P})$, then $\underline{P}(A \cap E) \leq P(A \cap E)$ and $P(\bar{A} \cap E) \leq \overline{P}(\bar{A} \cap E)$. Therefore,

$$P(A|E) \geq \frac{\underline{P}(A \cap E)}{\underline{P}(A \cap E) + \overline{P}(\bar{A} \cap E)}$$

and therefore $\underline{P}(A|E) \geq$ the same. The first part then follows from (14) and (12) by plugging in $\underline{P}(A \cap E) = \underline{P}(A||_*E)\underline{P}(E)$ and $\overline{P}(\bar{A} \cap E) = \overline{P}(\bar{A}||^*E)\overline{P}(E)$.

In the case where $\overline{P}(E) > \underline{P}(E) = 0$, if $\underline{P}(A||^*E) < 1$ the bound is trivially true. If $\underline{P}(A||^*E) = 1$ then $\overline{P}(\bar{A}||^*E) = 0 = \overline{P}(\bar{A} \cap E)/\overline{P}(E)$ so $\overline{P}(\bar{A} \cap E) = 0$ and $P(\bar{A}|E) = 0$ for any $P \in \mathcal{P}(\underline{P}(\cdot|E))$, so $P(A|E) = 1$.

Suppose $\underline{P}$ is 2-Monotone. Note that $(A \cap E) \cap (\bar{A} \cap E) = \emptyset$; therefore, there exists a $P \in \mathcal{P}(\underline{P})$ such that $P(A \cap E) = \underline{P}(A \cap E)$ and $P(\bar{A} \cap E) = \overline{P}(\bar{A} \cap E)$. Then (20) reduces to equality, as does (21) when $\underline{P}(A||^*E) < 1$. When $\underline{P}(A||^*E) = 1$, the tightness of (21) is trivially true. ■

The rule in (20) reduces to (11) for one-step conditioning, but as we discussed, (11) looses information and is thus problematic with incremental updates. The case when $\overline{P}(E) > \underline{P}(E) = 0$ was not considered by others such as [15] who have discussed (11), despite the fact that pragmatically it is important to know what to do in that case since it may in fact occur. Finally, it should be noted that when $\underline{P}$ is a probability distribution (and therefore also 2-monotone), the theorem reduces to the well-known Bayes' rule in probability theory.

It is also interesting to note from (23) that once an item of evidence is obtained such that $\underline{P}(E) = 0$, Dempster's rule contains, from that point on, all the information needed to compute the conditional lower probability.

5. Möbius Transforms $m(\cdot||_*E)$ and $m(\cdot||^*E)$

In this section we identify the Möbius Transforms of $\underline{P}(\cdot||_*E)$ and $\overline{P}(\cdot||^*E)$. We show that conditioning for these can be done directly in Möbius space, which is very convenient when the functions themselves are represented in the computer by their Möbius transforms. Although we do not use it here,

[24] has previously shown how to compute $\underline{P}(\cdot|E)$ as given in (11) directly in Möbius space.

THEOREM 1 (Dubois and Prade [13]) *Let $\underline{P}$ be a lower probability, m its Möbius transform, and $A, E \in \Omega$. Let $\underline{P}(A||_* E)$ be defined by (14). Then the Möbius transform of $\underline{P}(\cdot||_* E)$ is given by*

$$m(A||_* E) = \begin{cases} \frac{m(A)}{\underline{P}(E)} & \text{if } A \subseteq E \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

THEOREM 2 (Shafer [40]) *Let $\overline{P}$ be an upper probability function, m its Möbius transform, $A, E \in \Omega$. Let $\overline{P}(A||^* E)$ be defined by (12). Then the Möbius transform of $\underline{P}(\cdot||^* E)$ is given by*

$$m(A||^* E) = \frac{1}{\overline{P}(E)} \sum_{B:B\cap E=A} m(B) \tag{25}$$

Theorem 2 is known as Dempster's Rule of Conditioning. It is a well-known theorem (at least for the special case of belief functions) and is heavily used in the contexts of Dempster-Shafer theory ([40]) and the Transferable Belief Model ([50]).

The previous section showed that $\underline{P}$ can be updated incrementally, as can $\overline{P}$. It therefore follows that the Möbius transforms can also be updated incrementally as follows (provided they are defined, i.e., that $\underline{P}(E_1 \cap E_2) > 0$ or $\overline{P}(E_1 \cap E_2) > 0$ respectively):

$$m(A||^* E_1 \cap E_2) = \begin{cases} \frac{m(A||_* E_1)}{\underline{P}(E_2||_* E_1)} & \text{if } A \subseteq E_2 \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

$$m(A||^* E_1 \cap E_2) = \frac{1}{\overline{P}(E_2||^* E_1)} \sum_{B\cap E_2=A} m(B||^* E_1) \tag{27}$$

Implementing the update rules for $m(\cdot||_* E)$ and $m(\cdot||^* E)$ are quite easy. When a new item of evidence E is obtained, $m(\cdot||_* E)$ is obtained by throwing away any sets that are not totally contained within E and then normalizing the remaining Möbius assignments. Similarly, $m(\cdot||^* E)$ is computed by throwing away any Möbius assignments are not compatible with E, intersecting these with E, and then normalizing. Computing $\underline{P}(E)$ and $\overline{P}(E)$ each require a summation over focal elements as given by (3) and (4).

6. Example

This section demonstrates the use of the new representation using the example from Sections 2.1 and 3. The knowledge at any moment (after learning E) is represented by the following four items: $\langle m_*, m^*, p_*, p^* \rangle$, where $m_* = m(\cdot||_* E)$, $m^* = m(\cdot||^* E)$, $p_* = \underline{P}\,(E)$, and $p^* = \overline{P}\,(E)$. Initially, we can take $E = \Omega$, so that the initial knowledge is represented by: $\langle m_0, m_0, 1, 1 \rangle$, where m_0 is given in (5).

Suppose it is first learned that the true situation is in $E_1 = \{mammal, bird, reptile\}$. We compute $p_{1*} = \underline{P_1}\,(E_1) = \underline{P_0}\,(E_1||_*\Omega)\underline{P}\,(\Omega)$ by using (18), where $\underline{P_0}\,(E_1||_*\Omega)$ is obtained from m_0 using (3). In this example, this yields $p_{1*} = 0.04$. Similarly, we use (19) and (4) to compute $p_1^* = \overline{P_1}\,(E_1) = 0.5$. Finally, we use (26) and (27) to compute $m_{1*} = m(\cdot||_* E_1)$ and $m_1^* = m(\cdot||^* E_1)$ respectively. Our belief is now given by the four items: $\langle m_{1*}, m_1^*, 0.04, 0.5 \rangle$, where

$$m_{1*}(\{mammal\}) = 0.75 \qquad m_1^*(\{mammal\} = 0.06$$
$$m_1^*(\{reptile\}) = 0.92$$
$$m_{1*}(\{bird, reptile\}) = 0.25 \qquad m_1^*(\{bird, reptile\}) = 0.02$$

Next, suppose it is learned that the true situation is in $E_2 = \{mammal, bird, fish\}$. Precisely the same process is used, starting with $\langle m_{1*}, m_1^*, p_{1*}, p_1^* \rangle$. First, $p_{2*} = \underline{P_1}\,(E_2||_* E_1) \cdot p_{1*} = 0.75 \cdot 0.04 = 0.03$. Similarly, $p_2^* = \overline{P_1}\,(E_2||^* E_1) \cdot p_1^* = 0.08 \cdot 0.5 = 0.04$. The new updated belief is given by $\langle m_{2*}, m_2^*, 0.03, 0.04 \rangle$, where

$$m_{2*}(\{mammal\}) = 1 \qquad m_2^*(\{mammal\}) = 0.75$$
$$m_2^*(\{bird\}) = 0.25$$

Precisely the same result is obtained by conditioning the initial belief, $\langle m_0, m_0, 1, 1 \rangle$ with $E_1 \cap E_2 = \{mammal, bird\}$, or by performing the updates in the opposite order. With this representation, we can now compute bounds for any proposition of interest. For example, for the set $\{mammal\}$ using (22)

$$\underline{P}\,(\{mammal\}|E_1 \cap E_2) = \frac{1 \cdot 0.03}{1 \cdot 0.03 + 0.25 \cdot 0.04} = 0.75$$

7. Sparsity

Even if one were willing to ignore the loss of information from a straight lower-probability representation, there a still a complexity problem. Consider a lower probability function whose Möbius transform has k focal elements. After updating using

$$\underline{P}\,(A|E) = \frac{\underline{P}\,(A \cap E)}{\underline{P}\,(A \cap E) + \overline{P}\,(\bar{A} \cap E)} \tag{28}$$

the number of focal elements in the updated function may grow exponentially to $O(2^k)$ focal elements. On the other hand, the number of focal elements in the modified representation from the previous two sections actually never increases at all with new evidence. This may seem surprising since the modified representation captures more information than the straight representation, yet requires far fewer parameters to do so.

Ideally, the complexity of lower probabilistic inference should depend primarily on the number of constraints defining the function rather than on the size of the domain. In this way, we hope to take advantage of sparsity — in our case where the number of focal elements is small compared to $2^{|\Omega|}$.

The following example will demonstrate that an update with the straight lower probabilistic representation given by (28) can increase the number of focal elements exponentially. Recall that (28) produces tight bounds when $\underline{P}$ is 2-monotone.

Suppose $\underline{P}$ has k focal elements denoted by the sets $\mathcal{A} = \{A_1, A_2, ..., A_k\}$, such that any boolean combination of these sets is non-empty. An alternative way of stating this is to consider a surjective[4] mapping $X : \Omega \rightarrow \{0, 1, 2, ..., 2^k - 1\}$, and take A_1 to be all elements $\omega \in \Omega$ where the first bit is set when the results of $X(\omega)$ is written in binary, A_2 are those elements with the second bit set, and so on. Clearly we are considering a domain where $|\Omega| > 2^k$. For this example we can also assume that $\underline{P}$ is a belief function and that the Möbius Transform assignment for focal element $i = 1, .., k$ is $m(A_i) = 2^{i-1}/(2^k - 1)$. This $m(\cdot)$ has the property that the sum of Möbius assignments for each subset of $\mathcal{A}$ is unique.

Let $E = A_1 \cup A_2 \cup ... \cup A_{\frac{k}{2}}$. For simplicity, assume k is even. We will now count some of the focal elements in $\underline{P}\,(\cdot|E)$ — it is not necessary to count all of them because we can stop counting once we've accounted for an exponential number of them. As exact formula for obtaining the complete set of focal elements appears in [24, Section V].

[4] A surjective mapping is a function with the property that for any value in the range of the function, there is an input that maps to that value.

When $A \subseteq \Omega$, let

$$\begin{aligned} lc_E(A) &= \{A_i \in \mathcal{A} : A_i \subseteq A \cap E\} \\ uc_E(A) &= \{A_i \in \mathcal{A} : A_i \not\subseteq A \cup \bar{E}\} \end{aligned}$$

We can think of these as the *lower* and *upper core* of A in the expression

$$\underline{P}\,(A|E) = \frac{\underline{P}\,(A \cap E)}{\underline{P}\,(A \cap E) + \overline{P}\,(\bar{A} \cap E)}$$

In particular, $lc_E(A)$ and $uc_E(A)$ are the focal elements involved in the computation of $\underline{P}\,(A \cap E)$ and $\overline{P}\,(\bar{A} \cap E)$ respectively.

Suppose $A \subset A_1$, $A \neq A_1$. Then A is not a focal element of $\underline{P}\,(\cdot|E)$ because $\underline{P}\,(A \cap E) = \underline{P}\,(A)$ is zero for all such sets. However, $A = A_1$ will be a focal element, $lc_E(A_1) = \{A_1\}$, $uc_E(A_1) = \{A_2, ...A_k\}$, and

$$m(A_1|E) = \underline{P}\,(A_1|E) = \frac{\underline{P}\,(A_1 \cap E)}{\underline{P}\,(A_1 \cap E) + \overline{P}\,(\bar{A_1} \cap E)} = \frac{m(A_1)}{m(A_1) + \sum_{i=2}^{k} m(A_i)}$$

Next, consider supersets of A_1 that are not supersets of any of $A_2, A_3, ..., A_k$. The numerator in (28) for these sets will again be $m(A_1)$, but the second term of the denominator may change depending on the set. In fact, we can identify the sets on which the second term of the denominator change, which must therefore correspond to the addition of a focal element in the updated belief. Suppose, for example, that B contains A_1 and is only one element short of being equal to $A_1 \cup A_k$. Note that A_1 is in fact the only focal element contained by B. Then it also follows that $lc_E(B) = lc_E(A_1) = \{A_1\}$ and $uc_E(B) = uc_E(A_1) = \{A_2, .., A_k\}$. But $uc_E(A_1 \cup A_k) = \{A_2, .., A_{k-1}\}$, so that the computation of $\underline{P}\,(A_1 \cup A_k|E)$ involves a different upper core than the computation of $\underline{P}\,(B|E)$, a set that is only one element smaller. It must therefore be the case that $A_1 \cup A_k \cap E$ is a focal element in the updated function.

The same argument could have also been made for any of the sets $A_1 \cup A_{\frac{k}{2}+1}, A_1 \cup A_{\frac{k}{2}+2}, .., A_1 \cup A_{k-1}$, to identify a corresponding new focal element in the updated function. Furthermore, for each of these sets, the entire argument can be repeated further using their supersets to obtain additional focal elements. The result is as follows. Let $I \subseteq J = \{\frac{k}{2}+1, \frac{k}{2}+2, ..., k\}$, then the set $A_1 \cup \bigcup_{i \in I} A_i \cap E$ will be a focal element, with the lower probability given by

$$\underline{P}\,\left(A_1 \cup \bigcup_{i \in I} A_i \cap E|E\right) = \frac{m(A_1)}{m(A_1) + \sum_{A \in uc_E(A)} m(A)}$$

$$= \frac{m(A_1)}{m(A_1) + \left[\sum_{i=2}^{\frac{k}{2}} m(A_i) + \sum_{i \in J-I} m(A_i)\right]}$$

For each of the subsets of I, the upper core $uc_E(A_1 \cup \bigcup_{i \in I} A_i)$ is unique, and due to the previous choice of initial mass assignment, this implies that the updated lower probability is unique, and therefore there must be a non-zero mass-assignment to account for the difference.

Thus far we have identified a fraction of the sets that will be focal elements in $\underline{P}(\cdot|E)$. We have identified one focal element for each subset $I \subseteq \{\frac{k}{2}+1, \frac{k}{2}+2, ..., k\}$, making a total of $2^{k/2}$. Therefore, the number of focal elements in $\underline{P}(\cdot|E)$ is $O(2^k)$.

This example shows that the straight lower probability representation cannot, at least in the worst case, take advantage of sparsity (where the number of focal elements is much smaller than $|\Omega|$).

Next, consider the number of focal elements in the modified representation introduced in Sections 4 and 5. Recall from (24) that $\underline{P}(\cdot||_*E)$ is computed by deleting all focal elements of $\underline{P}(\cdot)$ except those contained within E, and then normalizing. Similarly, from (25), $\underline{P}(\cdot||^*E)$ is computed by deleting all focal elements from $\underline{P}(\cdot)$ except those that intersect E, intersecting the remaining ones with E and normalizing. Therefore, any focal element of $\underline{P}(\cdot||_*E)$ is also a focal element of $\underline{P}(\cdot||^*E)$. Thus, we need only count the focal elements of $\underline{P}(\cdot||^*E)$. If $\underline{P}(\cdot)$ has k focal elements, then $\underline{P}(\cdot||^*E)$ is guaranteed to have less than k focal elements. We see, therefore, that the number of focal elements in the new representation can only decrease as additional evidence is incorporated.

It is surprising that the straight lower probability representation is at the same time less informative and exponentially larger than our modified representation, but this is in fact what has just been shown. As a result, the modified representation is much more convenient to implement and use when reasoning about lower probabilities.

8. Conclusion

Lower probability has been used in many instances in existing literature in many different contexts and for many different reasons and purposes. It has often been suggested that these representations can be updated via conditioning in the same way as with exact probabilities by making use of interval arithmetic. However, it has been shown in this paper that there are a number of difficulties with doing so. Each update on such a representation looses information. The loss is not due to any particular conditioning rule, but occurs from a lack of representational power in the lower probability representation. With such a representation, the results of inference depend on what order evidence is incorporated, and on whether evidence is incorporated incrementally or all at once.

Computational considerations also present problems for the straight lower probability representation. It is not convenient to update the Möbius transforms of lower probability functions directly, and these representations cannot take advantage of sparsity (the case where the number of Möbius assignments defining the functions are small compared to the size of the domain). In fact, it was shown that the number of focal elements in a straight lower probability representation can increase exponentially when evidence is incorporated.

To rectify these problems, an alternative representation for lower probability was introduced. Rather than store the lower probability function directly, the new representation is composed of four different items, from which the lower probability of any set can be computed. Interestingly, one of these items is the function computed from Dempster's Rule of Conditioning, highlighting a new relationship between evidential reasoning and lower probability — namely, that Dempster's rule contains some, but not all, of the information needed to track lower probability.

Unlike with the straight lower probability representation, it is very convenient to directly update the Möbius transform of the new representation. The new representation does not experience the exponential growth with updates seen with the straight lower probability representation — in fact, the number of constraints (focal elements) actually gets smaller as more evidence is obtained. Furthermore, the new representation does not loose any information that is relevant for computing probability bounds. The same bounds are obtained regardless of what order evidence is incorporated, and regardless of whether updates are done incrementally or all at once.

The benefits and convenience of the new representation may be very helpful to anyone wishing to compute lower and upper probabilities. There are a number areas for future research. We have only considered one form of probabilistic inference in this document: conditioning. Other forms, for example Jeffrey's Rule, are possible ([7]) and we are currently experimenting with the use of similar rules on the new representation. It would also be interesting to further extend to connection between Dempster-Shafer theory and lower probabilities. In particular, since Dempster's Rule of Conditioning is actually a special case of Dempster's Rule of Combination, it would be interesting if a generalized rule of combination on the new representation could identified, and it would be interesting to examine if such a rule could enable a theory of statistical evidence in the spirit of [40, Chapter 9]. Finally, it would be very useful if methods for modularizing the new representation were developed, for example, in the spirit of [43], [45], and [3], and to develop local propagation methods as has been done for convex sets of probabilities in [4]. Such modularizations would be very significant, especially since they do not appear easy to come by.

Acknowledgement

The author was supported by the NASA-Jet Propulsion Lab under Grant No. NGT-51039 and the by NASA under grant No. NAGW-1175. This work greatly benefited from many discussions with my advisor, Reid Simmons, and from comments by Rich Goodwin, Sven Koenig, and Philippe Smets.

The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

References

1. J. Aitchison. Discussion on professor Dempster's paper. *Journal of the Royal Statistical Society, Series B.*, 30:234–237, 1968.
2. José Cano, Miguel Delgado, and Serafín Moral. Propagation of uncertainty in dependence graphs. In R. Kruse and P. Siegel, editors, *Proceedings of the European Conference ECSQAU on Symbolic and Quantitative Approaches to Uncertainty.*, pages 42–47, Marseille, France, October 1991. Springer-Verlag.
3. José Cano, Miguel Delgado, and Serafín Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *International Journal of Approximate Reasoning*, 8(4):253–180, 1993.
4. José Cano, Miguel Delgado, and Serafín Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *International Journal of Approximate Reasoning*, 8:253–280, 1993.
5. Alain Chateauneuf and Jean-Yves Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.
6. Gustave Choquet. Theory of capacities. *Ann.Inst.Fourier (Grenoble)*, 5:131–295, 1953.
7. Lonnie Chrisman. Abstract probabilistic modeling of action. In J. Hendler, editor, *First International Conference on AI Planning Systems*, College Park, MD, June 1992. Morgan Kaufmann.
8. G.F. Cooper. *NESTOR: A Computer-Based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge.* PhD thesis, Computer Science Department, Stanford University, November 1984. STAN-CS-84-48.

9. Luis M. de Campos, María T. Lamata, and Serafín Moral. The concept of conditional fuzzy measure. *International Journal of Intelligent Systems*, 5(3):237–246, 1990.
10. A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematics and Statistics*, 38:235–239, 1967.
11. A.P. Dempster. A generalization of Bayesian inference. *Journal Royal Statistical Society, Series B.*, 30:205–247, 1968.
12. Lorraine DeRobertis and J. A. Hartigan. Bayesian inference using intervals of measures. *The Annals of Statistics*, 9(2):235–244, 1981.
13. Didier Dubois and Henri Prade. On the unicity of dempster's rule of combination. *International Journal of Intelligent Systems*, 1(2):133–142, 1986.
14. Didier Dubois and Henri Prade. Evidence, knowledge, and belief functions. *International Journal of Approximate Reasoning*, 6(3):295–319, 1992.
15. Ronald Fagin and Joseph Y. Halpern. A new approach to updating beliefs. Technical Report RJ7222, IBM Almaden Research Center, Computer Science, 1989.
16. Jerome A. Feldman and Robert F. Sproull. Decision theory and artificial intelligence ii: The hungry monkey. *Cognitive Science*, 1:158–192, 1977.
17. Kenneth W. Fertig and John S. Breese. Probability intervals over influence diagrams. Technical report, Rockwell International Science Center, October 1991.
18. I. J. Good. Subjective probability as the measure of a non-measurable set. In *Logic, Methodology, and Philosophy of Science*, pages 319–329, Stanford, Cal., 1962. Stanford Univ. Press.
19. Ben Grosof. An inequality paradigm for probabilistic knowledge: The logic of conditional probability intervals. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 256–275. North Holland, 1986.
20. Joseph Y. Halpern and Ronald Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54(3):275–317, April 1992.
21. Eric J. Horvitz, John S. Breese, , and Max Henrion. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2:247–302, 1988.
22. Peter J. Huber. *Robust Statistics*. Wiley, New York, 1980.
23. Hidetomo Ichihashi and Hideo Tanaka. Jeffrey-like rules of conditioning for the dempster-shafer theory of evidence. *International Journal of Approximate Reasoning*, 3(2):143–156, 1989.
24. Jean-Yves Jaffray. Bayesian updating and belief functions. *IEEE Transactions on Systems, Man and Cybernetics*, 22(5):1144–1152, 1992.
25. Robert Kennes. Computational aspects of the Möbius transform of a graph. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(2):201–223, 1992.
26. Henry E. Kyburg, Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.
27. Henry E. Kyburg, Jr. Higher order probabilities and intervals. *Interna-*

tional Journal of Approximate Reasoning, 2:195–209, 1988.
28. Henry E. Kyburg, Jr. In defence of intervals. Technical Report TR268, University of Rochester, Computer Science, 1988.
29. Henry E. Kyburg, Jr. Knowledge. In J. F. Lemmer and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence, Vol. 2*, pages 263–272. Elsevier, Amsterdam, 1988.
30. Isaac Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, Mass., 1980.
31. Dennis V. Lindley. The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. *Statistical Science*, 2(1):17–24, 1987.
32. Serafín Moral and Luis M. de Campos. Updating uncertain information. In B. Bouchon-Meunier, R.R. Yager, and L.A. Zadeh, editors, *Uncertainty in Knowledge Bases. Third International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU*, pages 58–67. Springer-Verlag, 1991.
33. Richard E. Neapolitan and J. Kenevan. Justifying the principle of interval constraints. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, Mn., 1988. American Association for Artificial Intelligence.
34. Adrian Papamarcou and Terrence L. Fine. Unstable collectives and envelopes of probability measures. *The Annals of Probability*, 19(2):893–906, 1991.
35. Adrianos Papamarcou and Terrence L. Fine. A note on undominated lower probabilities. *The Annals of Probability*, 14(2):710–723, 1986.
36. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.
37. Judea Pearl. Reasoning with belief functions: An analysis of compatibility. *International Journal of Approximate Reasoning*, 4(5/6):363–389, 1990.
38. Enrique H. Ruspini. Approximate deduction in single evidential bodies. In *Proceedings of the Second Workshop on Uncertainty in AI*, pages 215–222, University of Pennsylvania, 1986.
39. Enrique H. Ruspini. Approximate inference and interval probabilities. In B. Bouchon and R.R. Yager, editors, *Uncertainty in Knowledge-Based Systems. International (1986) Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 85–94, Berlin, 1987. Springer-Verlag.
40. Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
41. Glenn Shafer. Constructive probability. *Synthese*, 48:1–60, 1981.
42. Glenn Shafer. Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4:323–362, 1990.
43. Glenn Shafer, Prakash P. Shenoy, and Khaled Mellouli. Propagating belief functions in qualitative markov trees. *International Journal of Approximate Reasoning*, 1:349–400, 1987.
44. Glenn Shafer and Amos Tversky. Languages and designs for probability judgement. *Cognitive Science*, 9:309–339, 1985.

45. Prakash P. Shenoy and Glenn Shafer. Axioms for probability and belief-function propagation. In R.D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. Elsevier Science Publishers B.V., North-Holland, 1990.
46. Philippe Smets. About updating. In B. D'Ambrosio, P. Smets, and P.P. Bonissone, editors, *Uncertainty in Artificial Intelligence*, pages 378–385, San Mateo, Ca, 1991. Morgan Kaufmann.
47. Philippe Smets. Probability of provability and belief functions. *Logique et Analyse*, 133–134:177–195, 1991.
48. Philippe Smets. The representation of quantified belief by belief functions: An axiomatic justification. Technical Report TR-IRIDIA-94-3.3, I.R.I.D.I.A. Université Libre de Bruxelles, Brussels, Belgium, 1994.
49. Philippe Smets. What is dempster-shafer's model? In R.R. Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theorey of Evidence*, pages 5–34. Wiley, New York, 1994.
50. Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.
51. Cedric A. B. Smith. Consistency in statistical inference and decision. *Journal Royal Statistical Society, Series B*, 23:1–25, 1961.
52. P. Snow. Bayesian inference without point estimates. In *National Conference on Artificial Intelligence (AAAI)*, pages 233–237, 1986.
53. David J. Spiegelhalter. Probabilistic expert systems in medicine: Practical issues in handling uncertainty. *Statistical Science*, 2(1):25–30, 1987.
54. Thomas M. Strat. Decision analysis using belief functions. *International Journal of Approximate Reasoning*, 4:319–417, 1990.
55. P. Suppes and M. Zanotti. On using random relations to generate upper and lower probabilities. *Synthesis*, 36:427–440, 1977.
56. Bj∅rnar Tessem. Interval probability propagation. *International Journal of Approximate Reasoning*, pages 95–120, 1992.
57. Carl Wagner and Bruce Tonn. Constructing lower probabilities. In D. Heckerman and A. Mamdani, editors, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 515–518, San Mateo, Ca., 1993. Morgan Kaufmann.
58. Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.
59. Peter Walley and Terrence L. Fine. Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, 10(3):741–761, 1982.
60. Larry A. Wasserman. Prior envelopes based on belief functions. *Annals Statistics*, 18:454–464, 1990.
61. Chelsea C. White. A posteriori representations based on linear inequality descriptions of a priori and conditional probabilities. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(4):570–573, 1986.