Combination Calculi for Uncertainty Reasoning: Representing Uncertainty Using Distributions

by

Robert Hummel[1]
Courant Institute of Mathematical Sciences, New York University

Larry M. Manevitz[2]
University of Haifa

[1]Courant Institute of Mathematical Sciences, New York University;

New York University
251 Mercer Street
New York, NY 10012

[2]Department of Mathematics and Computer Science, Univ. of Haifa;

Department of Mathematics and Computer Science
University of Haifa
Haifa 31905
Israel

Robert Hummel was a visiting professor at Vrije Universiteit, Amsterdam, The Netherlands, when the research for this work was conducted. He is currently on sabbatical at INRIA-Rocquencourt, Project Epidaure, BP 105, 78153 Le Chesnay cedex, France. Larry Manevitz is a frequent visitor to NYU's Courant Institute of Mathematical Sciences.

Combination Calculi for Uncertainty Reasoning: Representing Uncertainty Using Distributions

Robert Hummel and Larry Manevitz

Abstract

There are many different methods for incorporating notions of uncertainty in evidential reasoning. A common component to these methods is the use of additional values, other than conditional probabilities, to assert current degrees of belief and certainties in propositions. Beginning with the viewpoint that these values can be associated with statistics of multiple opinions in an evidential reasoning system, we categorize the choices that are available in updating and tracking these multiple opinions. In this way, we develop a matrix of different uncertainty calculi, some of which are standard, and others are new. The main contribution is to formalize a framework under which different methods for reasoning with uncertainty can be evaluated. As examples, we see that both the ''Kalman filtering'' approach and the ''Dempster-Shafer'' approach to reasoning with uncertainty can be interpreted within this framework of representing uncertainty by the statistics of multiple opinions.

1. Reasoning with uncertainty

Most expert systems make use of an evidential reasoning system, where evidence is combined with current belief states in order to maintain states of belief and confidence in a set of hypotheses. The fundamental concepts always involve quantities related to the degree of validity of a proposition, such as a probability, and other quantities related to the degree of certainty in the assertion of the degree of belief. Various calculi are used for representing these concepts and performing the calculations, including Bayesian networks, fuzzy logic, and the Dempster/Shafer theory of evidence. Each calculus has certain theoretical underpinnings, although a universally accepted methodology is still lacking.

It is now well-established that an evidential reasoning system must be able to deal with uncertainty. For example, the classical medical expert system MYCIN, and the famous geology expert system Prospector used different methods for handling degrees of certainty in propositions [1, 2]. Other significant systems that reason with uncertainty include INTERNIST [3], MUNIN [4], and INFERNO [5]. Degrees of certainty are represented both at the user-interface and within the automatic inferencing systems of these and most other successful expert systems, and one of the main problems in building an evidential reasoning system can be said to be the design of the method to handle uncertainty. Pearl's book offers a comprehensive survey of probabilistic reasoning and uncertainty methods [6]. An excellent introduction to the notion of representing uncertainty in probabilistic inferencing systems is given in Chapter 7 of Tanimoto's textbook [7] and in Neapolitan's book [8]. Collections of papers on uncertainty reasoning are found in a series of books compiled for the Workshops on Uncertainty in AI [9], and in the International Journal of Approximate Reasoning, for example.

There are really two problems that must be addressed when designing an evidential reasoning theory. First, the issues of belief, certainty, and confidence must be modeled in a rational manner. Second, the methodology for maintaining and combining states must be determined in a manner that conforms as nearly as possible to the model. One reason for the profusion of different calculi is that both issues present serious difficulties. The modeling issues present difficulties because different meanings can be ascribed to beliefs and certainties. Although probabilities are likely to be used to develop the calculus, the probabilities must apply to events that are well-defined, and the events will typically involve subjective evaluations that make the theory subject to varying interpretations. The methodological issues are difficult because no matter what scheme is chosen to implement the model, certain approximations will be necessary. Always, the methodology will fall short of the desired goals.

To make the issues more concrete, consider the difficulty of defining the statement that "This patient has a 20% probability of having disease D_1." The frequency interpretation of such a sentence means that among 100 patients having precisely the same symptoms and conditions, roughly 20 will have disease D_1. The difficulty with this interpretation is that it presupposes the existence of a sufficient number of cases with identical conditions — whereas the statement may be uttered by a knowledgeable physician who has never seen such a case; indeed, there may have never been such a case anywhere before! Another interpretation might be termed the "subjective" probability theory, and is founded on work by DeFinetti, Good, Savage, Kyburg, Fisher, and others [10-12]. In the "subjective interpretation," the statement can be interpreted to mean something along the lines of "I would accept 1 to 4 odds that the patient has disease D_1." However, such an interpretation can lead to different measurement methods. For example, if one insists that the bettor should come out even in the average over many bets on the same situation, then the subjective interpretation should yield the same value as computed in the frequency interpretation. On the other hand, one could take a particular situation and devise experiments to find a psychometric function under varying conditions, to find the odds under which a majority of experts would be indifferent when placing "bets." This value need not equal a precise frequency even if the frequency can reasonably be measured statistically.

Difficulties arise when inferencing is done using probabilities computed by the different methods. The frequency interpretation is supported by the availability of Bayes' theorem. Conditional probabilities can be computed using conjunctive and prior probabilities, and used to update and modify distributions. This, then, is ''Bayesian analysis.'' However, if probabilities are measured subjectively, then simple psychology experiments have shown that Bayes' formula is not obeyed by the values obtained from the psychometric functions [13]. Moreover, the average opinion need not equal the value obtained from a frequency analysis. Proponents of the ''Bayesian school'' ascribe this discrepancy to mistakes on the part of the subjects of the experiments. Others argue that Bayesian analysis is incompatible with the subjective interpretation.

Either way, there are problems in implementing a probabilistic reasoning system. If a purely Bayesian analysis approach is taken, then many different joint probabilities must be known in advance — generally more values than can be reasonably measured and stored. Some suggest that unknown values should be chosen by the principle of ''maximal entropy''; alternatively, independence assumptions can be invoked to give a functional relationship between groups of joint probabilities, essentially discarding certain conditional probabilities by saying that they are unimportant. More realistically, one can reason through intermediate hypotheses, with simple functional dependencies between the conditional probabilities from one node to the next, as in Pearl's system of networks [14]. The result is an algorithmic approach to modifying probabilities based on evidence accrued. One of the major appeals of Pearl's approach is that the functional dependencies are chosen to match expert rules. If an expert says that ''A causes B,'' the expert indicates not only a rule but many conditional independence relationships as well. Thus the resulting algorithm implements Bayesian updating according to natural causal relationships. However, the conditional independencies will be valid only to the extent that the experts correctly design the network.

If a non-Bayesian approach is taken, then any of a number of different calculi can be invoked, and considerable effort can be spent assessing the relative merits and the relationships among the various methodologies. Some of the potential methods include the Dempster/Shafer theory of evidence [15], or the use of fuzzy set theory [16]. However, once again, independence assumptions are generally necessary. More importantly, the semantics of the quantities in the representation must be interpreted subjectively.

In this paper, we approach the problem of modeling uncertainty in a somewhat different manner. We define our semantics *objectively*, by interpreting our measures of liklihood and uncertainty as objective, frequency measures of *opinions* about a certain situation. Thus, a 60% measure for the certainty in a diagnosis of measles given a certain set of symptoms, will no longer be interpreted as ''60% of patients having these symptoms would have measles''; but rather as saying ''Given this situation to evaluate, 60% of the experts would say the patient has measles.'' One can use other measures, but our common theme is that ''certainty'' is measured by the distribution of these opinions. That is, uncertainty is represented by a wide distribution of opinions while certainty is represented by close unanimity of opinions.

To use this idea, several issues should be addressed:

- What do we mean by an ''opinion'' of an expert?
- What information must be maintained? Is it necessary to maintain every separate opinion, or can some form of statistics be kept?
- What is the appropriate method for updating of the opinions? That is given new information, how should the opinions be revised?

We will assume that an ''opinion'' is an estimate of a quantity that is functionally related to a (frequency-based) probability, or is an estimate of a well-defined quantity representing the likelihood of a given proposition based upon given evidence. However, there is no need to assume consistency of opinions of a specific expert.

In the remainder of this paper, we present a variety of different calculi that are obtainable from specific choices, depending upon the values that the opinions are supposed to represent, and depending upon the assumptions used in the updating process. Our intent is to show the utility of the multiple-opinions approach to uncertainty in guiding the development of such calculi; we do not support a particular uncertainty calculus over all others. We do not purport to obviate other approaches to uncertainty. Some of these calculi are new; others turn out to be standard uncertainty calculi such as the Dempster/Shafer calculus, and the systems approach to combining uncertain estimates generally known as Kalman filtering. (As a side effect, this gives a Bayesian foundation to what are thought to be non-Bayesian combination formulae such as the Dempster/Shafer method.)

While we make no attempt to give a thorough survey of existing calculi nor their relation to the calculi that are derived in this paper (as a result of the multiple- opinions framework), we briefly review a small selection of alternative foundations to compare them with our formulation.

This uniform presentation should be useful in order to *choose* a calculus appropriate for a specific application. That is, we make clear how various different assumptions result in the different calculii.

2. Uncertainty and distributions of opinions

The use of intervals of probabilities in order to represent uncertainty in reasoning is an appealing idea. While a Bayesian system, or other probability-based system, normally uses a single probability in order to represent the chance or likelihood in a particular proposition, frequently a degree of belief in that value is also required. One method of representing the degree of certainty, discussed early on by Dempster [17] and others [10, 18], is to use an interval of probabilities, say [10%,90%] to indicate little certainty in a 50% probability estimate, or [45%,55%] to indicate relative certainty in the central 50% probability. However, the use of the Dempster combination formula in conjunction with the interpretation of the belief and plausibility functions as a probability interval has led to illogical behavior [19]. Alternative methods tag the indicated probability with a ''certainty value,'' which might be interpreted as a scaled version of the probability window, and thus can form an equivalent representation.

A good way to view the probability interval formulation of the uncertainty representation is to distinguish between two sample spaces. Consider, as an example, the case of medical diagnosis of a patient. There is the sample space of all patients, from which can be determined prior and conditional probabilities for various diseases based on various symptoms. However, there can be a second sample space, the sample space of doctors, with each doctor giving an opinion or estimate of a specific probability. The fact that the doctors' opinions are subjective estimates is not incompatible with the existence of true probabilities: since there are two sample spaces, there can be different functions. If all of the doctors agree on a particular probability, and providing the sample space is large enough, then there is considerable confidence, or belief, in that probability. However, if the doctors give a wide range of opinions, then we have an interval of probabilities, which might be determined from the minimum and maximum opinion, or might be represented by a standard deviation spread in the opinions from the mean. In either case, it is clear that we can view a range of probabilities as a confidence measure in a particular probability, by assuming that the range arises from multiple subjective estimates.

Figure 1 depicts the situation for the medical diagnosis situation, illustrating the different sample spaces that occur: namely, the spaces of patients, of experts, of diseases, and of symptoms.

Since the sample space of patients is distinct from the sample space of doctors, the opinions of the doctors need not be based on a statistically significant number of cases, and thus the existence of probabilities over the space of patients may be problematical. Nonetheless, the doctors are able to estimate values, given precise instructions, related to likelihoods based on physiological (i.e., causative) analyses, or philosophical predictions of probabilities.

We can refer to the doctor's estimates as probabilistic values, or functionally related to probabilistic values, even though there may be no evidence for the existence of probabilities. Alternatively, the doctors' opinions might be, for example, fuzzy values; what matters is that the values should be defined in some precise way, so that combinations of pairs of values are well-defined.

Accordingly, we posit that uncertainty in a probabilistic value can be represented by a collection of estimates of the quantity, and that the degree of certainty or uncertainty can be measured by means of the distribution of values in the collection of estimates. The collection of estimates, or opinions, forms a random process over a sample space that we will call the set of "experts" E. It is important to realize that all of the experts are attempting to estimate the same quantity based on the same information. That is, every expert has the same set of evidence (which is what we mean by a collection of experts). If new evidence becomes available, then every expert will revise his or her opinion, or perhaps the set of experts will be replaced with a new set of experts. This is in

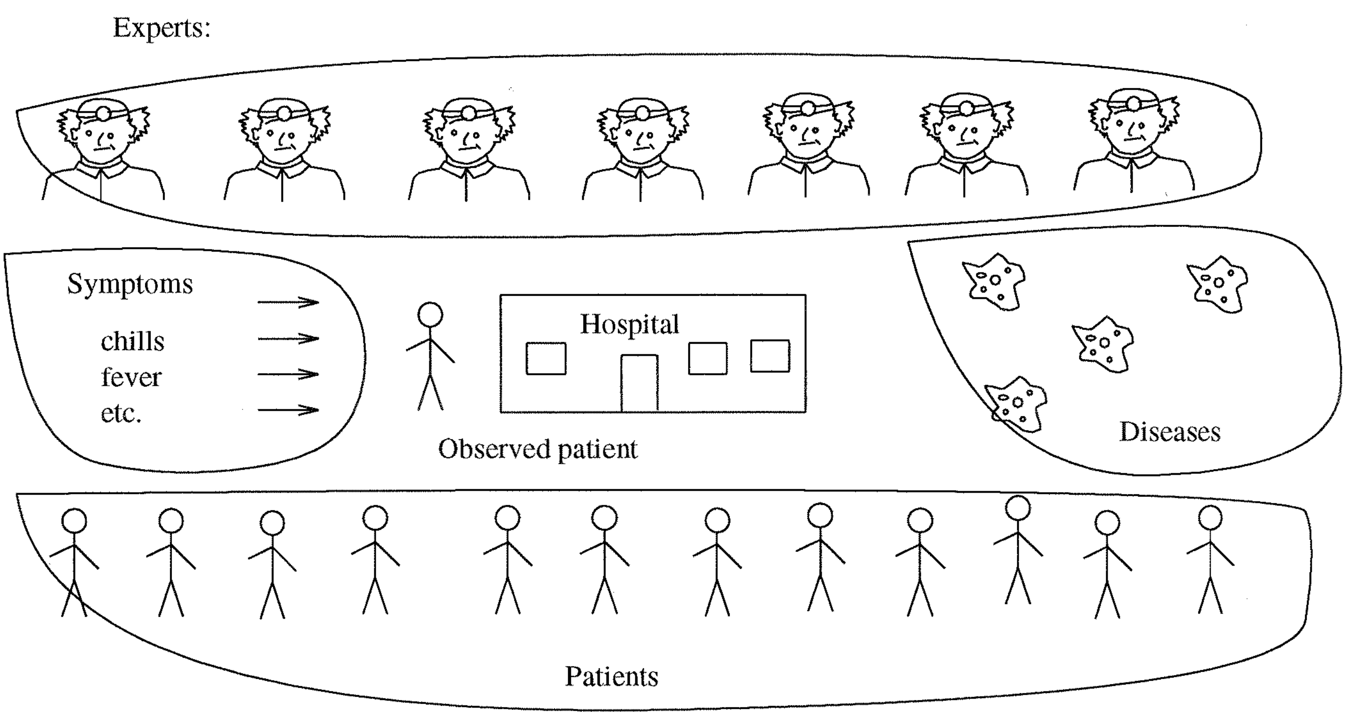

Figure 1. The various sample spaces involved in our medical diagnosis application. Probabilities are defined in terms of frequencies over the space of all patients. Certainty is measured in terms of the distribution of opinions over the space of doctors, after all the evidence is combined.

distinction to schemes where each piece of evidence yields a different conditional probability, and one examines the distribution of those probabilities over the set of different pieces of evidence [20].

In our representation of uncertainty, the cause of the uncertainty is not the variation or the discordance of the evidence. The uncertainty that we wish to model is the intrinsic uncertainty that arises from a particular piece of evidence. Thus the collection of experts share one piece of information, or one set of information, and nonetheless have a spread of opinions. This spread comes from a lack of certainty, or the fact that estimates must be made, or perhaps due to a lack of a statistically viable base of experience. When we combine the opinions of the experts with another set of experts having different opinions related to different (or new) information, the new level of uncertainty arises purely due to the new spread of opinions generated by the combination method. Instead, one might look at the different pieces of evidence, and estimate levels of uncertainty not according to the discordance of opinions, but based on the unlikelihood of the particular mix of evidence.

For example, different pieces of information may give rise to different levels of uncertainty; we might call this extrinsic uncertainty, because it comes from an analysis of the mix of different pieces of information. One piece of information, in isolation, might argue strongly for the diagnosis of a particular disease, say D_1 with a high degree of certainty. Another piece of evidence, in isolation, argues for the same disease, D_1, again with high certainty. The two pieces of evidence, in concert, might argue for disease D_1, but with a high degree of **un**certainty, because the disease often occurs with one symptom or the other, but the co-occurrence of both is associated with many possibilities. In this case, the extrinsic information about the combination of evidence leads to a large increase in uncertainty, simply due to the confluence of the two pieces of information. On the other hand, an intrinsic combination of the two will likely result in a high degree of certainty for disease d_1, since both constituent opinions were certain. Another possibility is that the two pieces of evidence argue for different diseases in isolation, with moderate uncertainty, but in confluence argue for a third disease with high certainty. Computing extrinsic uncertainty depends entirely on the particulars of the situation, and can make use of a model of causality of the diseases. However, it is clear that no general statement can be made about the functional dependence of the certainties on the range of evidence.

As we will discuss in Section 5, in the absence of other assumptions, the probability distribution over a collection of labels, conditioned on two different pieces of information, can have an arbitrarily complex functional relationship depending on the probabilities for each label, conditioned on each piece of information individually.

Similarly, the level of uncertainty for each label, given two pieces of information, may be arbitrarily related to probabilities *and* uncertainties for those labels conditioned on each piece of information individually. To make order of the chaos, one must either have a functional model for the relationships, or one must make certain simplifying assumptions. For the case of the probability distributions, the requisite assumptions typically involve independence of the information; for the uncertainties, we will assume the lack of extrinsic uncertainty.

More general treatments that can account for extrinsic uncertainty are required by certain applications. Different methods could be proposed for dealing with the required modeling of the functional relationships. The methods used could generalize the calculi considered here, or alternative computational systems, such as Bayesian networks, might be used to model the updating of uncertainty due to extrinsic information (see [8] Chapter 10,[6] Chapter 7, and [21]).

3. Related theories of uncertainty

3.1. The Dempster/Shafer Calculus

The well-known Dempster/Shafer ''theory of evidence'' [15] can be formulated precisely as a representation of multiple subjective opinions of propositions [22]. In this formulation, it can be seen that the combination rule invokes a Bayesian updating between pairs of opinions, but that since there are many opinions encoded in the state, the overall updating formulas are not Bayesian. However, the theory of evidence makes use only of boolean opinions [22], where the ''doctors'' give only a list of possibilities, as opposed to assigning probabilities to each proposition. Indeed, this is how Dempster originally formulated the states used for his combination formula [17]. The subset-valued function $\Gamma(\omega)$, used in Dempster's formulation, forms the set of propositions that the sample (i.e., expert) ω says are possible, given the information currently available. Different samples ω can give rise to different opinions of the set of possibilities $\Gamma(\omega)$. Dempster proceeds to give a method for combining multiple sets of opinions in a quite general framework, and this formulation has become the basis of the theory of belief functions underlying the Dempster/Shafer theory of evidence. Let us review briefly the relationship between the belief function and the distribution of opinions of $\Gamma(\omega)$, and the combination method, in terms of opinions of doctors in a medical diagnosis application.

For a given patient, the set of possible diseases is given by Λ. Each doctor $\omega \in E$ states the subset of possible diseases $\Gamma(\omega) \subset \Lambda$ that he cannot rule out. All doctors have the same information, but not all doctors have the same opinion (i.e., the same list of possible diseases). The mass $m(A)$ on a particular subset $A \subset \Lambda$ is the percentage of doctors that state that the subset of possible diseases is precisely A. The belief $Bel(A)$ on a subset A is the percentage of doctors that have ruled out all diseases outside of A, i.e., for whom $\Gamma(\omega) \subset A$. Finally, the plausibility $Pl(A)$ is defined as the percentage of doctors whose subset of possible diseases intersects A. [1]

New evidence is represented by another collection of doctors E', each with his own list of opinions, $\Gamma'(\omega')$ for $\omega' \in E'$. A new state is formed as follows (see also Figure 2). A new set of "doctors" is formed, consisting of the set of all committees of two doctors, with one doctor from the original set E and one doctor from the new set E'. Each of these composite doctors $(\omega, \omega') \in E \times E'$ forms a new opinion by taking the intersection of the opinions of the constituent doctors, i.e., $\Gamma(\omega) \cap \Gamma'(\omega')$. Thus a committee rules out all diseases that are ruled out by either of the members of the committee. However, if the resulting committee has ruled out all the diseases in Λ, then that committee is discounted, and is dropped from the collection of new doctors. Assuming that not all committees drop out, it can then be shown that the percentage of the remaining committees that state that A is precisely the subset of possible diseases is given by:

$$\frac{\sum_{B \cap C = A} m(B) \cdot m'(C)}{1 - \sum_{B \cap C = \varnothing} m(B) \cdot m'(C)},$$

which is precisely the Dempster rule of combination. More details of this formulation are given in [22].

The Dempster/Shafer theory requires two kinds of independence. Because intersections are used in combining opinions, there is an implicit assumption of independence of the information being combined. The evidence does not have to be independent in the true probabilistic sense, but only in a sense of possibilities. The exact requirement is that if a hypothesis is ruled out by a particular symptom, then it is must be ruled out by that symptom in combination with any other symptom. (See the discussion of independence-of-evidence assumptions in Pearl's rejoinder to discussions on [19], available as [23].) In addition, because a set product is used to form the new collection of experts, each expert is being paired independently with all other experts in the other set. This is a second independence assumption. Both independence assumptions limit the general utility of the calculus. Another difficulty is that all opinions are boolean — doctors only give lists of possible diseases, and do not offer any

[1] It is interesting to compare the interpretation implied by our formulation of belief functions with Pearl's useful and successful interpretation [19] based on conditional probabilities of provability. The interpretations are compatible; where we use a collection of doctors, Pearl speaks of a collection of logical theories from which the hypothesis can be proved. The isomorphism is only broken by allowing the doctors to make estimates.

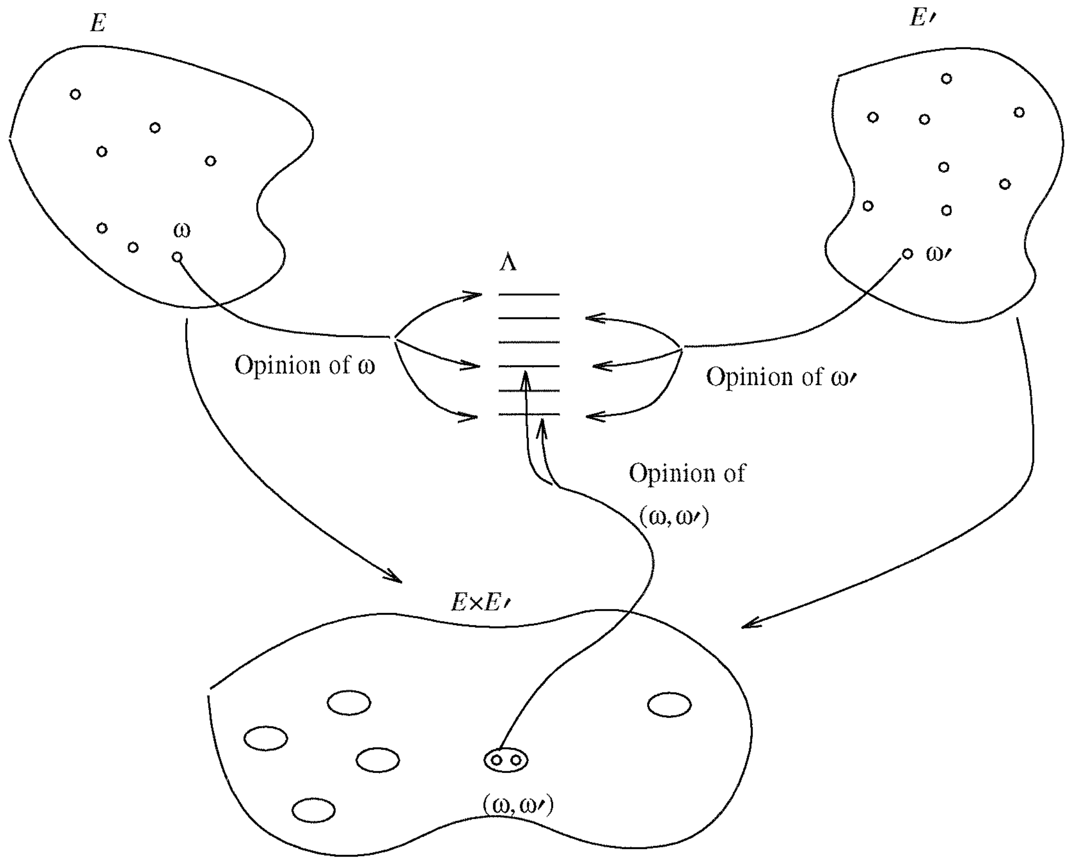

Figure 2. Combination in the Dempster/Shafer theory of evidence. Opinions are represented by subsets of Λ. A state is given by the statistics of the opinions over a set of experts *E*. A pair of states is combined by computing the statistics of the opinions over the product set of experts, where pairs of opinions are formed by intersecting component opinions.

estimates of the likelihoods or probabilities of the diseases.

3.2. Yen's GERTIS system

There have been many attempts to relax the objection that the Dempster/Shafer calculus is based purely on possibilities, despite the appearance of continuous values that look like probabilities. For example, Yen [20] extends Γ so that each element maps to a collection of disjoint nonempty subsets together with a probability distribution over the collection of subsets. That is, $\Gamma^*(\omega)$ is a probability distribution over a collection of (disjoint) subsets, rather than giving a single subset. Yen then develops a calculus with a modified Dempster's formula. However, for Yen, each element in E has a different body of evidence. The masses are still computed by computing statistics over E. Accordingly, this methodology is applicable when the elements of information in E are largely independent, and should be combined statistically rather than inferentially. In combination with another collection of information E', the information is paired off and combined using conditional independence assumptions between elements of E and elements of E'. The principal difference between this approach and the approach taken here is that here we assume that all elements of E have the same information, and that variation occurs because each element in E maps to an estimate, rather than an exact measurement. We also do not attempt to extend our approach to the hierarchical label sets that Yen uses.

3.3. Pearl's incorporation of uncertainty in Bayesian nets

Pearl's comprehensive treatment of probabilistic reasoning [6] includes a study of uncertainty representation and propagation in probabilistic reasoning systems such as in Bayesian networks.

Bayesian networks, called by Neapolitan ''independence networks,'' and also called ''causal networks,'' typically use directed acyclic graphs with random variables associated with each node and conditional probabilities associated with each edge. While there are extensions of the methods described below to general directed acyclic graphs, for simplicity we discuss the case of singly connected graphs. (See, for example, Lauritzen and Spiegelhalter [24] for a discussion of one form of Bayesian networks for multiply-connected directed acyclic graphs.) An underlying assumption for the singly-connected graph is that any pair of nodes are conditionally independent given the information at a node that lies on the directed path between them. This is a very natural assumption for expert systems, for example, since a graph can be created from a system of ''causal rules''. Additional independence relations are implied by the causal network. The independence assumptions relate naturally to intuitive causality rules: given the immediate antecedent of a conclusion, the conclusion is independent of non-immediate causes. So it can be argued that the expert rules developed by the human experts reflect the indepedence relations as understood by the expert in the system.

Under these assumptions, Pearl and his associates have shown that there is a direct algorithmic relationship that allows the propagation of new information through the network. The propogation methods are derived from the conditional independence assumptions and Bayes rule.

Pearl has also suggested [6] (Section 7.3) that one can use this algorithmic relationship to carry out a sensitivity analysis, and thus to develop an uncertainty combination formulation. Specifically, one could use probability intervals for the information at nodes as a means of representing uncertainty, and propagate all possible combinations of probabilities within the interval. Alternatively, one could use a probability distribution over a range of values at one or more nodes, and compute the expected state of the network. By defining certain nodes to be inputs (actually, the information given for the node), and other nodes to be outputs, a Bayesian network can be used to define a functional combination method. This is a logical way to define a potentially complex combination formula. Further, the formula can be extended to deal with a representation of uncertainty, using the sensitivity computation that deduces a density distribution over the outputs given distributions for the inputs. Thus, a network can be used to develop a functional updating formula, including uncertainty. We make use of functional updating formulae when developing our calculi, although typically we consider far simpler updating methods.

In Section 7, we comment on possible augmentations to Bayesian networks in light of the theories developed here.

3.4. Tzeng's mathematical formulation of uncertainty

Tzeng has introduced a mathematical model of uncertain information [25]. The model is related to the Dempster/Shafer calculus, in that there are ''messages'' that map to subsets of labels, but is considerably more general. Indeed, Tzeng shows how the Bayesian calculus and the Dempster/Shafer calculus fit into the scheme, and demonstrate a range of other possibilities that can be derived from the same model.

There is a close relationship between his model and our viewpoint of multiple opinions of experts; the main difference occurs in the kinds of opinions that are permitted. Whereas our nominative notion of opinions will be a probability distribution over the labels, Tzeng, as with the Dempster/Shafer theory of evidence, consider only boolean opinions (that is, lists of possible labels). However, each message in Tzeng's formulation may have a weight (see our space of weighted experts in Section 4.3), and for a given piece of evidence s (a code, in Tzeng's terminology), there is an associated collection of distinct messages (i.e., experts) $E_1(s)$, $E_2(s)$, etc., where we recall that each message gives a boolean opinion. From our viewpoint, this collection is a set of opinions, based upon the information s. Then, for different pieces of evidence s, one obtains different such collections (although note that there is an implicit ordering: that $E_i(s_1)$ and $E_i(s_2)$ are in some sense related — see our notion of pairwise combining of opinions in Section 4.3.4). The collection S of possible codes s is then given a prior probability distribution, and computations may be made in a Bayesian fashion conditioned on certain codes or certain messages occurring. We see, however, that Tzeng's information space, which is the full collection of codes, messages, and mappings of codes to messages and messages to subsets of labels, in some sense represents the collection of all possible computational domains for a given labeling situation, confining the various pieces of possible evidence to give rise to collections of boolean opinions (and thus Belief function representations). As we discussed in the previous section, the functional relationship between the possible pieces of evidence and the probabilities and uncertainties is in general unrestricted, and thus any function over an information space might be a reasonable model for a particular situation. The modeling that leads to structure in Tzeng's model occurs when the variety of available messages, codes, and sets of experts, are restricted.

The model that we develop differs from Tzeng's formulation primarily in the way that choices lead to different calculi; our formulation requires that the choices be stated in advance, and influence the way that maps and spaces come together.

4. The Models

In this section, we present the alternatives for the models that will, in combination, lead to different calculi for uncertain reasoning. Each model, as described above, depends on states that are based on multiple non-boolean opinions, where each opinion is expressed relative to a set of propositions (i.e., hypotheses). We will say that each opinion is offered by an expert and that:

E represents the set of experts,

and that:

Λ represents the set of propositions (or labels).

For simplicity, we will assume that $\Lambda = \{1,2, \cdots ,N\}$. We will refer to these as labels, although they are variously referred to as the "frame of discernment," the "outcomes," or the "hypotheses."[2] The important point is that the labels form a mutually exclusive and exhaustive set of possibilities: exactly one label is true for any given situation.

[2] Hypotheses typically have binary truth values; thus n output hypotheses can be viewed as a method of representing 2^n true output possibilities (labels) through a binary encoding of each label.

For each expert $\omega \in E$, the map

$$x_\omega : \Lambda \rightarrow \mathbb{R}$$

gives ω's opinions for each of the labels $\lambda \in \Lambda$. Since we have chosen to represent the labels by integers, the opinions of expert ω forms a vector, $\mathbf{x}_\omega = (x_\omega(1), x_\omega(2), \cdots x_\omega(N))$. Mathematically, we can regard $\mathbf{x}_\omega$ as a realization of a random vector **x**, representing sample ω from the sample space E.

A calculus for uncertain reasoning will result when we explain how to mix two groups of opinions. That is, we will assume that we have two sets of experts: E_1 and E_2. Each set of experts have their opinions, respectively $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, containing the opinions $\mathbf{x}^{(1)}_{\omega_1}$ and $\mathbf{x}^{(2)}_{\omega_2}$ for $\omega_1 \in E_1$, $\omega_2 \in E_2$. What is desired is a method to determine a single combined set of experts E, each with their opinions, $\mathbf{x}_\omega$ for $\omega \in E$.

The choices, each of which yields a different calculi, relate to:

- The values represented by the values in the opinions (i.e., the $x_\omega(\lambda)$);
- Which statistics are used to track (and simplify) the opinions;
- The way in which the two sets are combined (e.g., by set product or by set union); and
- The formulas used to combine the opinions of a single pair of experts.

We emphasize that the statistics used to describe the distribution of opinions forms the representation of uncertainty, thus obviating the necessity to maintain individual expert's opinions. The practical feasibility of implementing an uncertainty calculus using the representation is related to the complexity of the statistical parameters that are used to describe the distribution.

Let us consider the options.

4.1. Values represented by the opinions

We describe three alternatives for the representation of opinions: **probabilities, log probabilities**, and **odds**.

4.1.1. Probabilities. The most obvious choice is to allow each expert to give an estimate of a **probability** for each label:

$$x_\omega(\lambda) = \text{Probability of } \lambda \text{ according to } \omega.$$

If we denote the information available to the experts by s, then each expert is making an estimate of $\text{Prob}(\lambda | s)$. The resulting opinion, $\mathbf{x}_\omega$, is thus the estimate of a probability vector. Clearly,

$$0 \leq x_\omega(\lambda) \leq 1.$$

It would seem desirable that each expert should specify a valid probability vector, so that

$$\sum_{\lambda=1}^{N} x_\omega(\lambda) = 1. \qquad \text{(Tentative)}$$

However, we will not impose such a requirement on our experts. There is no reason to impose such a requirement, since the experts are making estimates, and not defining precise probabilities. In fact, the estimates may well be of subjective probabilities. However, each estimate is based on the given evidence, and is made independently of the estimates for the alternative labels. Thus if we point out to an expert that his estimates are inconsistent because the sum of his probabilities is either less or more than one, he should not be perturbed, and merely answer that his estimates are not exact.

We might have instead required our experts to constantly normalize their estimates so that the sum is always one. In practice, this is the most likely scenario. However, for technical reasons, we will not insist that experts perform normalizations. There are two problems with normalizations. The first is that some aberrant expert might give estimates of 0 for all labels, and thus give a vector that can not be normalized. However, this problem can be circumvented by banning the zero-vector estimate, or perhaps by banning all zero estimates — opinions must lie in the range $0 < x_\omega(\lambda) \leq 1$. However, in these cases, it is important to then use an updating formula which ensures that the restrictions are respected after combination. The second, more serious difficulty has to do with the fact that the average of a set of normalized probability vectors may not be the same as the normalization of the average of the unnormalized opinions. Thus if we track only statistics of the set of experts, updating by the experts will not factor through to formulas on the statistics if we insist on normalizations. We will further see, when we discuss updating using independence assumptions, that we avoid certain difficulties if we permit our experts to give estimates that are not necessarily normalized.

4.1.2. Log probabilities. A disadvantage of the probability opinions is that many updating schemes will require knowledge of the prior probabilities of each label. That is, updating using probability opinions will be parameterized by a collection of constants, $\gamma_1, \cdots, \gamma_N$, one for each label, where $\gamma_\lambda = \text{Prob}(\lambda)$ is the prior probability for label λ.

As an alternative, we will consider the **log-probability** opinions, defined as:

$$x_\omega(\lambda) = \log\left[\frac{\text{Probability of } \lambda \text{ according to } \omega \text{ based on the evidence}}{\text{Prior probability of } \lambda \text{ as estimated by } \omega}\right].$$

That is, each expert, having information s, makes an estimate of the quantity

$$L(\lambda|s) = \log\left[\frac{\text{Prob}(\lambda|s)}{\text{Prob}(\lambda)}\right].$$

The value $x_\omega(\lambda)$ is expert ω's estimate of $L(\lambda|s)$. Note that this value is a so-called "log-likelihood," in that it is equal to $\log(\text{Prob}(s|\lambda)/\text{Prob}(s))$, by Bayes' theorem. The log-probability $L(\lambda|s)$ is a real number, positive or negative, that indicates the influence of the information s on the probability of the label λ in terms of orders of magnitude (base e). Thus, for example, an estimate $x_\omega(\lambda) \approx -2$ says that the information s decreases the probability of label λ by roughly two orders of magnitude (i.e., by a factor of e^{-2}) relative to the prior probability. This method of representation, without the uncertainty calculus, is described by Charniak and McDermott [26].

The opinion vector $\mathbf{x}_\omega$ in this case can have positive and negative components, and encodes the information of the conditional probabilities. Indeed, if the prior probabilities $\text{Prob}(\lambda)$ are known, then the conditional probabilities $\text{Prob}(\lambda|s)$ can be recovered from the set of log-probabilities, $L(\lambda|s)$. The precise formula is

$$\text{Prob}(\lambda|s) = \frac{\text{Prob}(\lambda)\cdot\exp(L(\lambda|s))}{\sum_{\lambda'=1}^{N} \text{Prob}(\lambda')\cdot\exp(L(\lambda'|s))}.$$

Similarly, an *estimate* of the conditional probabilities can be obtained from the *estimates* of the log-probabilities in the log-probability opinion vector $\mathbf{x}_\omega$ using the same formula, substituting $x_\omega(\lambda)$ for $L(\lambda|s)$. Because of the normalization in the above formula, an additive constant to all components of $\mathbf{x}_\omega$ is unimportant: $\mathbf{x}_\omega$ represents the same opinion as $\mathbf{x}_\omega + (c,c,\cdots,c)$. However, if each expert ω uses a different $\mathbf{c} = (c,c,\cdots,c)$, then this will affect the estimate of the uncertainty by skewing the covariance matrix. Thus we require that all experts use the same vector $\mathbf{c} = 0$ when making their log-probability estimates. However, the fact that the representation is invariant to translations in this way is useful for the accumulation of evidence in the absence of certain independence assumptions.

4.1.3. Odds formulation. Yet another alternative representation of $x_\omega(\lambda)$ concerns the special case when the label set Λ has only two alternatives. In this case, we typically have that $\Lambda = \{\text{H}, \neg\text{H}\}$, where H is some proposition. That is, the label set consists of a proposition and its negation, and the two relevant probabilities are simply Prob(H) and Prob($\neg$H). In this case, each expert needs to maintain only one value, and a reasonable form of the representation is to use x_ω as a vector of length one, i.e., a scalar, to estimate the odds:

$$O(\mathrm{H}) = \frac{\mathrm{Prob(H)}}{\mathrm{Prob}(\neg \mathrm{H})}.$$

Both probabilities are recoverable from the odds by the formulas Prob(H) = O(H)/(1+O(H)), and Prob(¬H) = 1 – Prob(H). Using $x_\omega \approx O(\mathrm{H})$ results in the **odds formulation**. The odds values are always nonnegative, but are unbounded. In this case, there is an implicit normalization, since only one parameter is tracked, although there are two constituent probabilities that could be estimated.

4.2. Statistics

Next, we consider the statistics that should be maintained by the uncertainty reasoning system. There are many possibilities, but we will list only three. We consider the **complete representation**, the tracking of **mean and covariance** statistics, and **mean and standard deviations** statistics. We omit higher order tracking, and methods that might model, say, the distribution as a sum of multinormal distributions.

4.2.1. Complete representation. The **complete representation** keeps all the values, and tracks all the opinions. Generally, however, such a system involves too many values. If the opinions record discrete values, such as boolean values stating whether a label is possible or impossible, then a complete set of joint probabilities can be maintained. This is exactly the situation in the Dempster/Shafer theory. The representation in this theory is based on a belief function, which is updated by combining with other belief functions. The belief function encodes all of the joint statistics over the labels [22], where by joint statistics we mean the probabilities over the set of experts. Thus the belief in a subset A is the probability that an expert names a subset of possibilities that is contained in A. The mass on A is the joint probability that the labels in A are precisely the ones that are named as being possible by a random expert (taken from the sample of experts). The joint probability, that a particular collection A of labels are named (with no restrictions on the other labels) is known as the commonality value $Q(A)$ of the set. From the collection of mass values, or plausibilities values, or commonality values, it is possible to derive the other collections; thus each collection is an equivalent representation of the full 2^N set of joint probabilities. In essence, rather than maintaining N probabilities, one for each label in Λ, the Dempster/Shafer calculus uses 2^N values, one for each subset of Λ. It is important to recall, however, that these statistics are computed over the sample space of experts, and not over the sample space of problem instances. Once again, however, the Dempster/Shafer formulation generally involves unacceptably many variables. Various methods have been proposed for reducing or simplifying the number of variables in the Dempster/Shafer calculus [27, 28].

For the Dempster/Shafer calculus, each opinion is given by N boolean values, and thus there are 2^N possible opinions. In general, when there are P discrete possible opinions, then the complete set of statistics can be specified by 2^P values. When the opinions are values from a continuous range, and therefore not limited to a finite set, then the full set of statistics will generally require storing each and every opinion. Accordingly, when the opinions are values from a continuous domain, then the complete statistics representation simply amounts to tracking each and every opinion.

4.2.2. Mean and Covariance. Instead, if we regard map $x_\omega : \Lambda \to \mathbb{R}$ as a realization of a random vector over the sample space $\omega \in E$, then the most obvious statistics to maintain are the mean and covariance values, defined by:

$$\mu(\lambda) = \operatorname*{Avg}_{\omega \in E} \left[x_\omega(\lambda) \right]$$
$$C(\lambda,\lambda') = \operatorname*{Avg}_{\omega \in E} \left[\left[x_\omega(\lambda) - \mu(\lambda) \right] \cdot \left[x_\omega(\lambda') - \mu(\lambda') \right] \right].$$

Since the covariance matrix is symmetric, there are a total of $N + N(N+1)/2$ variables in the **mean and covariance** representation, given that there are N labels.

4.2.3. Mean and standard deviations. As an even simpler alternative, we might retain only the N means and the N standard deviations. The standard deviation for label λ, $\sigma(\lambda)$, can be defined as:

$$\sigma(\lambda) = \sqrt{C(\lambda,\lambda)}.$$

This is the **mean and standard deviations** representation. Whereas the mean and covariance representation models the distribution of opinions as an ellipsoidal cloud about the mean opinion, the mean and standard deviation representation uses a squashed spherical shape, i.e., an ellipsoid oriented along all axes, about the mean in an N-dimension Euclidean ''opinion space'' (see Figure 3).

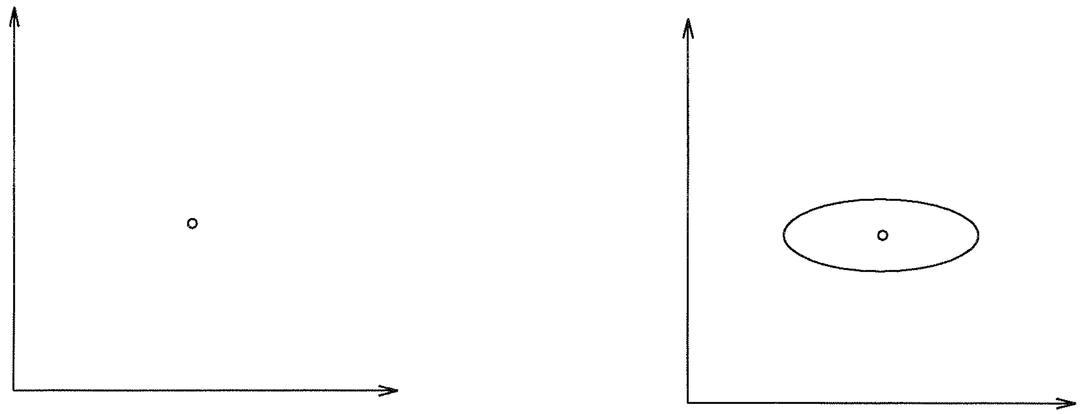
a b

Figure 3. A multinormal spread of opinions by (a) the mean and covariance representation and (b) by the mean and standard deviation representation.

In some of the calculi, we need in addition to the above statistics a value for the total number of experts in E, representing a total weight M_E for the given piece of evidence shared by the experts in E.

So far, we have described the set-up assuming that each expert has equal weight. We can easily extend the representation to allow each expert to have a nonnegative weight $\rho(\omega)$, insisting that the sum of weights $\sum_{\omega \in E} \rho(\omega) = M_E$ equals the total weight (or mass) of the evidence associated with the experts E, if a total weight value is included. When the experts have different weights, the calculation of the average in the definition of the mean and covariance values should take the constituent weights into account. When this is done, the mean is given by

$$\mu(\lambda) = \frac{1}{M_E} \sum_{\omega \in E} \rho(\omega) \cdot x_\omega(\lambda),$$

and the covariance is

$$C(\lambda, \lambda') = \frac{1}{M_E} \sum_{\omega \in E} \rho(\omega) \cdot \left[x_\omega(\lambda) - \mu(\lambda) \right] \left[x_\omega(\lambda') - \mu(\lambda') \right].$$

For both the "mean and covariance" and "mean and standard deviations" representations, the narrowness of the distribution is a measure of the certainty in the mean opinion. There are various ways of reducing the multinormal distribution to a single measure of the degree of certainty. For example, we can take the volume of the ellipsoid, which can be computed from the determinant of the covariance matrix, as a measure of uncertainty. Alternatively, we can measure uncertainty using the maximum semi-diameter of the ellipsoid, or, depending on the values represented by the opinions, we might want to measure maximum *relative* variation from the mean value within the ellipsoid, where the relative variation of each component is measure as a percentage of the component value in the mean opinion.

4.3. Combining sets

We now turn to the alternatives for the formulation of the combination method. In this subsection, we are concerned with the method for creating a single set of experts E from two given sets E_1 and E_2. We will define the methods of combining by **union**, by **pairwise matching**, by **set product** and by the **product measure**. In addition, we introduce the **non-commutative update by means** and **commutative update by means** methods.

4.3.1. Set union. The most obvious way is simply to take a union:

$$E = E_1 \cup E_2.$$

The **union** method of combination, which seems so desirable, has a serious flaw: namely, evidence is never actually combined, but only pooled. It is the essential feature of evidential reasoning that one body of information known to E_1 in combination with another body of information E_2 can lead to a conclusion that would not be known or suspected by either group individually. The union method has no way of incorporating such connectives into the reasoning system. Thus the union method of combining experts leads to formulas that do not depend on the method of combining opinions, since pairs of opinions are never combined.

4.3.2. Set product. The **product method** of combining the sets is to replace the union with a set product:

$$E = E_1 \times E_2,$$

so that E becomes the set of all committees of pairs of experts, composed of one expert from each of the two sets, E_1 and E_2. This is, for example, the method used in the Dempster/Shafer calculus. The advantage of this method is that each committee can use an opinion combination method (discussed in the next subsection) to reason about the combined information.

4.3.3. Product measure. In the case when each expert has a weight, we need to define the weights of the committees. In this case, each expert ω_1 in E_1 has a weight $\rho_1(\omega_1)$, and each expert ω_2 in E_2 has a weight $\rho_2(\omega_2)$, so that in the product set formulation, the resulting weight of the committee $\omega = (\omega_1, \omega_2)$ should be

$$\rho(\omega) = \rho_1(\omega_1)\cdot\rho_2(\omega_2).$$

The total weight M_E of the product set E is simply the product of the constituent total weights, $M_{E_1}\cdot M_{E_2}$. Since the weights on the experts forms a measure, this representation is simply the **product measure** of experts.

4.3.4. Pairwise matching. The **pairwise** method assumes that the number of experts is always indexed over the same set, say $\{1 \cdots n\}$. Thus we may assume that the experts are numbered 1 through n, in any given set of experts E. Then given two sets of experts, E_1 and E_2, each set containing n ordered experts, then the combined set E is also n ordered experts, where the k^{th} expert is formed by combining the k^{th} expert in E_1 with the k^{th} expert in E_2. That is, the collection of experts is regarded as an n-tuple, and combinations are performed pairwise on the tuples. The length n is fixed in advance, and can be made as large as we like.

In fact, in order to get meaningful formulas, we will generally assume that n is infinite, so that the experts are indexed over the natural integers $\{1,2 \cdots\}$, and that the errors (i.e., variations from the mean) in the opinions are always orthogonal. Thus if $\{x_i^{(1)}(\lambda)]>_{i=1}^{\infty}$ is one collection of opinions for any given label, with mean $\mu^{(1)}(\lambda)$, and $\{x_i^{(2)}(\lambda')\}_{i=1}^{\infty}$ is another collection of opinions (perhaps for a different label) with mean $\mu^{(2)}(\lambda')$, then the expectation

$$E\{(x_i^{(1)}(\lambda)-\mu^{(1)}(\lambda))\cdot(x_i^{(2)}(\lambda')-\mu^{(2)}(\lambda'))\} = 0.$$

This orthogonality condition on the noise, however, holds only between pairs of collections of opinions. The condition is implied by independence of the opinions $x_i^{(1)}(\lambda)$ and $x_i^{(2)}(\lambda')$ viewed as random variables over the sample space of natural integers, but is weaker than complete independence. In the case when the opinions are gaussian random variables, then orthogonality of the errors is equivalent to independence.

For the case of updating using probabilities, a further assumption will be needed, which is equivalent to complete independence. Thus if we consider the opinions x_ω to be random vectors over the sample space of experts, then we require independence of the collection of random vectors obtained from different sets of experts (which in the pairwise matching case are indexed over the same set). Independence can be viewed roughly as saying that the joint cumulative probability distribution functions are separable into a product of the constituent probability functions [29], but practically, for our purposes, will mean that

$$E\{(x_i^{(1)})^k\cdot(x_i^{(2)})^l\} = (\mu^{(1)})^k\cdot(\mu^{(2)})^l$$

for any pair of sets of opinions on either the same or different labels.

Interestingly, these independence assumptions are required only for the pairwise matching combination method, whereas in the product method, the independence of cross terms comes about by explicitly including the terms.

4.3.5. Non-commutative update by means. There are two other alternatives that lie somewhere between the union method and the product method of combining sets of experts. In the first, the **non-commutative update by means**, each expert in E_2 combines with the mean opinion of the experts in E_1, potentially taking into account the spread of opinions in E_1. The result is a collection of opinions E, with one opinion for each member of E_2. This method is not commutative, since it matters whether experts in E_2 combine with the mean opinion of experts in E_1, or experts in E_1 combine with the mean opinion in E_2.

4.3.6. Commutative update by means. The **commutative update by means** method combines each expert in E_1 with the mean of the opinions in E_2, and each expert in E_2 combines with the mean of the opinions in E_1, to form a collection of opinions indexed over E, one for each expert in the union $E_1 \cup E_2$.

4.4. Combining opinions

In nearly all of the models, the core of the evidential reasoning process occurs when two opinions must be combined. Each opinion is an estimate of a quantity related to a probability, so we discuss methods for updating probabilities. Our viewpoint is purely Bayesian, even though the eventual formulas for evidential reasoning among multiple opinions will not follow Bayes' law. We will cover in detail two Bayesian-based combination methods, one where **conditional independence** is assumed, and the other where we introduce **α-dependence** assumptions. In addition, we discuss **conjunctive** updating, and **functional** updating.

4.4.1. Conditional independence updating. Let us suppose that two experts are to combine their information. The first expert knows the information s_1, and the other expert knows the information s_2. Their opinions are based on estimates of quantities related to $\text{Prob}(\lambda|s_1)$ and $\text{Prob}(\lambda|s_2)$ respectively, for all $\lambda \in \Lambda$. Of course, what is desired is knowledge of the distribution $\text{Prob}(\lambda|s_1,s_2)$ for $\lambda \in \Lambda$.

If we assume **conditional independence** of the evidence, i.e.,

$$\text{Prob}(s_2|s_1,\lambda) = \text{Prob}(s_2|\lambda), \ \ \forall \lambda$$

then there is an easily-defined updating function. In that case, it is not hard to show that

$$\text{Prob}(\lambda|s_1,s_2) = \frac{1}{K} \cdot \frac{\text{Prob}(\lambda|s_1) \cdot \text{Prob}(\lambda|s_2)}{\text{Prob}(\lambda)}$$

where K is a normalization constant, and is given in two forms by

$$K = \frac{\text{Prob}(s_1, s_2)}{\text{Prob}(s_1) \cdot \text{Prob}(s_2)} = \sum_{\lambda' \in \Lambda} \frac{\text{Prob}(\lambda' | s_1) \cdot \text{Prob}(\lambda' | s_2)}{\text{Prob}(\lambda')}.$$

The computation is straightforward and standard, and proceeds from Bayes' Law and the fact that conditional independence is equivalent to the statement that

$$\text{Prob}(s_1, s_2 | \lambda) = \text{Prob}(s_1 | \lambda) \cdot \text{Prob}(s_2 | \lambda).$$

Note that conditional independence is symmetric, so that the resulting combination formula is commutative: updating information s_1 with information s_2 is the same as updating s_2 with s_1.

It is also worth noting, since it is a common error, that conditional independence neither implies nor is implied by unconditional independence: $\text{Prob}(s_1, s_2) = \text{Prob}(s_1) \cdot \text{Prob}(s_2)$. If we assert unconditional independence in addition to conditional independence, then we can additionally conclude that $K = 1$. However, this can lead to a host of problems, since it puts compatibility constraints on the constituent probability vectors, and hence on the $\text{Prob}(\lambda | s_1)$'s and $\text{Prob}(\lambda | s_2)$'s. Specifically, suppose we are given $\text{Prob}(\lambda | s_1)$'s, and assume conditional independence and unconditional independence. Then in order to combine these probabilities with a set of $\text{Prob}(\lambda | s_2)$'s, we must have that the resulting $\text{Prob}(\lambda | s_1, s_2)$'s, given by the above combination formula with K=1 is a valid probability distribution. This constraint means, however, that only certain $\text{Prob}(\lambda | s_2)$'s can be allowed: there will be an additional constraining equation, and the constraint will depend on the probabilities given from information s_1. Thus the assumptions of conditional independence together with unconditional independence imposes restrictions on sets of allowable probability vectors, which is hardly realistic in practice. In a certain sense, the assumptions of conditional independence is irreconcilable with an assumption of unconditional independence.

These problems disappear when we use *opinions* for the updating. This is because the opinions are viewed as estimates, and are not required to equal probability vectors. In the case when the opinions represent estimates of probabilities, we do not require that the estimates sum to one, and so a pair of estimates $\mathbf{x}^{(1)}_{\omega_1}$ and $\mathbf{x}^{(2)}_{\omega_2}$, obtained using information s_1 and s_2 respectively, may update under the assumption of conditional and unconditional independence according to the formula $x_{(\omega_1, \omega_2)}(\lambda) = x^{(1)}_{\omega_1}(\lambda) \cdot x^{(2)}_{\omega_2}(\lambda) / \gamma_\lambda$, where γ_λ is the prior probability $\text{Prob}(\lambda)$. The fact that the resulting vector may not sum to one does not concern us, since the combined opinion from the committee (ω_1, ω_2) is only an estimate. In this case, both conditional and unconditional independence may be used, and there is no incompatibility. For updating probabilistic opinions, we will assume both types of independence (the assumption of unconditional independence will be dropped for updating of log probabilities).

4.4.2. Alpha dependence updating. Independence is extremely useful, but often quite unrealistic. It requires many separate and strong statements about the probabilities, and implies that the symptoms have no underlying common mechanism. Especially in reasoning networks, conditional independence may not be valid. One method for modeling some level of dependence is introduced in [30], where the authors have introduced the notion of **α-dependence**. The assumption required is that there exists a real value $\alpha(s_1,s_2)$ such that the following N conditions hold:

$$\text{Prob}(s_2|s_1,\lambda) = \Big[\text{Prob}(s_2|\lambda)\Big]^{\alpha(s_1,s_2)}, \quad \lambda = 1,2, \cdots N.$$

The essential requirement here is that $\alpha(s_1,s_2)$ is independent of λ. Note that if $\alpha = 1$, then we have the same assumption as conditional independence. The case $\alpha = 0$ might be called complete dependence; that is, the information s_2 is implied by the information s_1. We may also have $\alpha > 1$, which signifies an inverse dependence: information s_1 makes s_2 less likely in the presence of any given label, always by the same exponential amount.

Under the assumption of α-dependence, for $\alpha = \alpha(s_1,s_2)$, we have the updating formula

$$\text{Prob}(\lambda|s_1,s_2) = \frac{\dfrac{\text{Prob}(\lambda|s_1)\cdot[\text{Prob}(\lambda|s_2)]^{\alpha}}{\text{Prob}(\lambda)^{\alpha}}}{\sum\limits_{\lambda' \in \Lambda} \dfrac{\text{Prob}(\lambda'|s_1)\cdot[\text{Prob}(\lambda'|s_2)]^{\alpha}}{\text{Prob}(\lambda')^{\alpha}}}.$$

The denominator of these expressions is the same as

$$K = \frac{[\text{Prob}(s_1,s_2)]}{\text{Prob}(s_1)\cdot[\text{Prob}(s_2)]^{\alpha}},$$

which will be equal to one if we additionally assume *unconditional α-dependence*:

$$\text{Prob}(s_1|s_2) = [\text{Prob}(s_1)]^{\alpha}.$$

However, we will only require both conditional and unconditional α-dependence when updating probabilistic opinions. An advantage of the log-probability representation is that the formulas work fine even if K is not equal to one.

It is interesting to note that when $\alpha = 0$, then the addition of information s_2 adds no new information, and so no updating takes place. That is, the resulting distribution is the same as the initial distribution. Also, symmetry is not guaranteed. The fact that s_2 is $\alpha(s_1,s_2)$-dependent on s_1 does not imply that there will exist a coefficient $\alpha(s_2,s_1)$ for an α-dependence relationship for s_1 on s_2, except in the case $\alpha = 1$. Even if an $\alpha(s_2,s_1)$-relationship does exist in addition to the given $\alpha(s_1,s_2)$, the relationship is likely to exist only approximately, and the optimal values for the two α's are not functionally constrained. Generally, given information s_1 and s_2 from respective classes of information, an α-relationship will only exist in one of the two directions, and then only approximately. Accordingly, given distributions for $\text{Prob}(\lambda|s_1)$ and $\text{Prob}(\lambda|s_2)$, the system should combine the two using either an $\alpha(s_1,s_2)$ formula, or an $\alpha(s_2,s_1)$ formula, whichever is appropriate. In typical usage, we have a current body of knowledge, represented by s_1, and some new evidence, s_2, and we wish to combine the new knowledge with the old. Let us assume that the evidence is accumulated in an order such that when an α-relationship exists, it is the $\alpha(s_1,s_2)$ coefficient that is known, so that the new information s_2 influences the existing information with an exponent of $\alpha(s_1,s_2)$ as in the formula above. For noncommutative updating by means, and for the product updating methods, this is all that is needed.

For the commutative updating by means, we must assume that both an $\alpha(s_1,s_2)$ coefficient and a $\alpha(s_2,s_1)$ coefficient are known, and model the α-dependence of the data s_2 on s_1 and the data s_1 on s_2. In order not to obtain constraints on the allowable probability distributions given information s_1 and s_2, it is necessary to assert that the α-dependence conditions will hold only approximately. In this case, $\alpha(s_1,s_2)$ will most likely represent an average over λ of the ratio of the logarithms of the probabilities, and similarly for $\alpha(s_2,s_1)$.

The updating formulas for the conditional independence and the α-independence cases become particularly simple when viewed in terms of the log-probabilities. The formula for the α-independence case is

$$L(\lambda|s_1,s_2) = L(\lambda|s_1) + \alpha(s_1,s_2)\cdot L(\lambda|s_2),$$

and the independent case is given simply by $\alpha = 1$. Recall that additive constants to all components of L are unimportant, so that we can ignore the normalization factor in the denominator of the equation for $\text{Prob}(\lambda|s_1,s_2)$, since it is independent of λ.

Updating of the odds given α-dependence depends upon the prior odds of the proposition P. If we define the prior odds to be

$$O(\text{H}) = \frac{\text{Prob(H)}}{\text{Prob}(\neg\,\text{H})},$$

then the conditional odds update according to

$$O(\text{H}|s_1,s_2) = O(\text{H}|s_1)\cdot\left[\frac{1}{O(\text{H})}\cdot O(\text{H}|s_2)\right]^{\alpha(s_1,s_2)}.$$

The independent case is again given by $\alpha = 1$. Recall that in the case of the odds representation, only the one value, the conditional odds on the proposition H, forms the entire opinion vector of any given expert ω. The quantity in the brackets in this formula, $O(\mathrm{H}|s_2)/O(\mathrm{H})$ is the same as the likelihood-ratio: $\mathrm{Prob}(s_2|\mathrm{H})/\mathrm{Prob}(s_2|\neg\mathrm{H})$.

4.4.3. Conjunctive updating. We next consider the case of **conjunctive** combination of opinions. In this formulation, in a pair of opinions, the most pessimistic opinion dominates. This translates to the formula

$$\mathrm{Prob}(\lambda|s_1,s_2) = \frac{1}{K}\min\{\mathrm{Prob}(\lambda|s_1),\mathrm{Prob}(\lambda|s_2)\}.$$

Here K is a normalizing constant that is needed in order to ensure that the left hand side is a valid probability distribution. In practice, the minimum is taken of the two opinions $x_{\omega_1}(\lambda)$ and $x_{\omega_2}(\lambda)$, and the result is assumed to be monotonically related to the probabilities. That is, the normalization is ignored, and the collection of values comprising any given opinion $x_\omega(1), \cdots, x_\omega(N)$ does not even approximate a probability distribution. Since different results would ensue if the normalization were performed for every combination, it might be the case that the values of the opinions (the "fuzzy values") are not functionally related to probabilities. Heckerman has a detailed discussion of the use of the 'min' operator in MYCIN [31] and its relationship to probabilities. The conjunctive combination rule is an example of a fuzzy logic rule [16, 32] which is nonetheless applicable in some situations. A conjunctive combination is most common in the two-label case, and applies only to one of the two labels, and thus will be considered only in terms of the "odds formulation" for the representation of the opinions. For this case, the conjunction formula is

$$O(\mathrm{H}|s_1,s_2) = \min\{O(\mathrm{H}|s_1),O(\mathrm{H}|s_2)\}.$$

The renormalization necessary to maintain an underlying probability distribution happens automatically by allowing the "not H" proposition to absorb discarded probability weight. Generalized combination rules for other logical connectives of propositions can be handled in a similar manner. Bayesian networks, for example, use "noisy-or gates," ([6], Section 4.3.2) and other methods to model kinds of dependence in a related fashion.

4.4.4. Functional updating. In general, anything is possible for the combined distribution. Knowledge of $\mathrm{Prob}(\lambda|s_1)$ and $\mathrm{Prob}(\lambda|s_2)$ in no way constrains $\mathrm{Prob}(\lambda|s_1,s_2)$. Each of the N outputs is an arbitrary function of the $2N$ input values, subject only to the conditions that the inputs and outputs are probability distributions. We can write

$$\mathbf{r} = \boldsymbol{\pi}_{s_1,s_2}(\mathbf{p},\mathbf{q}),$$

where $\mathbf{p}$ and $\mathbf{q}$ are the probability vectors with components $\mathrm{Prob}(\lambda|s_i)$, λ=1, $\cdots$,N for i=1,2 respectively, and $\mathbf{r}$ is the resulting probability vector formed from the components $\mathrm{Prob}(\lambda|s_i,s_j)$. Depending on the information s_1,s_1,

$\boldsymbol{\pi}_{s_1,s_2}$ may be known, unknown, or known imprecisely. It commonly occurs that $\boldsymbol{\pi}_{s_1,s_2}$ is only known reasonably well in a subregion of its domain. Under certain assumptions, $\boldsymbol{\pi}_{s_1,s_2}$ is known over the entire domain and has a simple form. In the functional updating case, it is assumed that a model for $\boldsymbol{\pi}_{s_1,s_2}$ is constructed for every instance of a pair of information sources (s_1, s_2).

Each of the previous cases constitute an assumption on the form of the function $\boldsymbol{\pi}_{s_1,s_2}$. For example, in terms of $\boldsymbol{\pi}_{s_1,s_2}$, conditional independence asserts that

$$\boldsymbol{\pi}_{s_1,s_2}(\mathbf{p},\mathbf{q}) = \frac{1}{\sum_{j=1}^{N} \frac{p_j \cdot q_j}{\gamma_j}} \cdot \left(\frac{p_1 q_1}{\gamma_1}, \frac{p_2 q_2}{\gamma_2}, \ \cdots, \frac{p_N q_N}{\gamma_N} \right)$$

where the γ_j are a vector formed from the prior probabilities Prob(λ) for $\lambda = 1 \cdots N$. Likewise, conjunctive updating gives a simple formula for $\boldsymbol{\pi}_{s_1,s_2}$. One can design other examples of functionals $\boldsymbol{\pi}_{s_1,s_2}$ that might be applicable in certain cases.

4.5. Summary of the options

Table 1 contains a summary of the options that can be used in developing the models. Each column contains the alternatives for a particular component of the system, and the choice for each column is substantially independent of the other columns. That is, in order to build a model for evidential reasoning with uncertainty, the designer should choose one from column A, one from column B, etc.

5. The tables of uncertainty calculi

We proceed to describe the calculus for each reasoning model that can be obtained from the table. We will treat sequentially the three possibilities for the form of representation of the opinions, namely as probabilities, log-probabilities, and the odds formulation. We will only consider the mean and covariance statistics — the case of mean and standard deviation statistics can generally be obtained as a special case of the mean and covariance. The case of complete statistics, for our formulations with continuous-valued opinions, is accomplished simply by tracking all opinions, and thus is omitted. Within each heading, we have choices for the combination of experts and the method for combining opinions, leading to a two-dimensional table. Finally, we develop a somewhat different calculus based on Kalman filtering in the next section. The updating formula involves a product instead of a sum in

Representation of Opinions	Statistics to retain	Combining of Sets of Experts	Combining of Opinions
Probabilities	Complete	Union	Conditional independence
Log-probabilities	Mean & covariance	Pairwise matching	α-dependence
Odds formulation	Mean & std deviation	Product	Conjunctive
		Product measure	Functional
		Noncommutative by means	
		Commutative by means	

Table 1. Summary of the options for the models.

these cases.

We reiterate that the Dempster/Shafer calculus fits into the scheme, but with different choices: namely, using a boolean representation for the opinions. In that case, the Dempster/Shafer calculus can be shown to arise when all statistics are tracked (not just the first-order means and second-order covariances, but all order statistics up to N, the number of labels), assuming a product set combination of sets of experts and independent updating of experts. However, the principal limitations, as mentioned earlier, are that boolean opinions are less interesting than probabilistic ones, and that the full set of statistics involves 2^N variables, whereas the simpler means and covariance statistics use only $O(N^2)$ variables.

To describe the combination formulas, we assume that we have two states: the current knowledge is accounted for by the state $\mathbf{x}^{(1)}$ having means $\mu_1(\lambda)$ and covariance matrix $C_1(\lambda,\lambda')$; the state $\mathbf{x}^{(2)}$ to be combined with the current state has means $\mu_2(\lambda)$ and covariance matrix $C_2(\lambda,\lambda')$. In addition, when needed, we assume that the total weights of the respective states are M_1 and M_2.

For simplicity, we denote by α_1 the value for $\alpha(s_1,s_2)$, and by α_2 the value for $\alpha(s_2,s_1)$. Throughout, we use $\boldsymbol{\mu}$, $\boldsymbol{\mu}_1$, and $\boldsymbol{\mu}_2$ to denote the vectors with components $\mu(\lambda)$, $\mu_1(\lambda)$, and $\mu_2(\lambda)$, and $\mathbf{C}$, $\mathbf{C}_1$, and $\mathbf{C}_2$ to denote the matrices with entries $C(\lambda,\lambda')$, $C_1(\lambda,\lambda')$, and $C_2(\lambda,\lambda')$. Also, we use the notation $\mathbf{x}^2$ for the vector outer product of an N-vector with itself: $\mathbf{x}\cdot\mathbf{x}^T$. This is an N by N matrix whose λ,λ' component is $x(\lambda)\cdot x(\lambda')$. The termwise product of two matrices, A and B, is here denoted by AB, and yields a matrix whose λ,λ' component is $A(\lambda,\lambda')\cdot B(\lambda,\lambda')$. The termwise product of two vectors is denoted by $\mathbf{x}.\mathbf{y}$, and yields a vector with components $x(\lambda)\cdot y(\lambda)$. The vector

$\boldsymbol{\alpha}$ stands for the N-vector all of whose components are α, and similarly for $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ and $\mathbf{-1}$, and a vector raised to a vector $\mathbf{x}^{\mathbf{y}}$ is just termwise exponentiation; in particular, $\boldsymbol{\gamma}^{-1}$ has components $1/\gamma_\lambda$ for $\lambda = 1, \cdots, N$. The values m_1 and m_2 are the masses M_1 and M_2 normalized so that $m_1 + m_2 = 1$. A table of all abbreviations is given in Table 2.

We now make some initial global comments on the formulas.

The union method of updating leads to the same formula, regardless of the method of combining opinions, as noted before, since there is no pairwise combination of opinions that takes place. In fact, the formulas for the revised mean and covariance when the union method is used are the same formulas for the probabilities, log-probabilities, and the odds form of representation of the opinions.

Abbreviated Notation	Actually Denotes
α_1, α	$\alpha(s_1, s_2)$
α_2	$\alpha(s_2, s_1)$
$\boldsymbol{\mu}$	$(\mu(1), \mu(2) \cdots \mu(N))$
$\boldsymbol{\mu}_i$	$(\mu_i(1), \mu_i(2), \cdots \mu_i(N))$
$\mathbf{C}$	$\left[C(\lambda, \lambda')\right]_{\lambda,\lambda'}$
$\mathbf{C}_i$	$\left[C_i(\lambda, \lambda')\right]_{\lambda,\lambda'}$
m_1, m_2	$\frac{M_1}{M_1+M_2}, \frac{M_2}{M_1+M_2}$
$\mathbf{x}^2$	$\mathbf{x}\cdot\mathbf{x}^T = \left[x(i)\cdot x(j)\right]_{i,j}$
AB	$\left[A(\lambda, \lambda')\cdot B(\lambda, \lambda')\right]_{\lambda,\lambda'}$
$\mathbf{x}.\mathbf{y}$	$(x(1)y(1), \cdots, x(N)y(N))$
$\boldsymbol{\alpha}$	$(\alpha, \alpha, \cdots, \alpha)$
$\mathbf{x}^{\mathbf{y}}$	$(x(1)^{y(1)}, \cdots, x(N)^{y(N)})$
$\boldsymbol{\gamma}^{-1}$	$(1/\gamma_1, \cdots, 1/\gamma_N)$
$\phi(\boldsymbol{\mu}_1, \mathbf{C}_1, \boldsymbol{\mu}_2, \mathbf{C}_2)$	$(\boldsymbol{\mu}, C)$ of min of two Gaussian random variables (Formulas (5.1) and (5.2) below)
$\psi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \mathbf{C}_2)$	$(\boldsymbol{\mu}, C)$ of min of constant $\boldsymbol{\mu}_1$ with Gaussian random variable (Formula (5.3) below)

Table 2. Abbreviations used in the formulas.

The product measure method for combining experts leads to the same formulas as the product method. The fact that the two sets of experts might have unequal weights has no effect on the formulas. This is a consequence of the fact that we endow a committee of two experts (ω_1, ω_2) with the product of their respective weights, and that we normalize the averages by dividing out by the product of the total weights $M_1 \cdot M_2$.

The pairwise matching formulation, with the appropriate independence assumptions, leads to the same formulas as the product methods. This occurs because the independence assumptions are precisely the ones needed to disregard the ''cross terms'' in the combinations of experts that the product method includes explicitly. The pairwise method of combination is the more common assumption in stochastic modeling and systems approaches to dealing with measurement inaccuracies, as with Kalman filtering. The product method, on the other hand, is the foundation of the Dempster rule of combination, and of other uncertainty calculi. Our results, thus, point out the distinction between the approaches, but unify the methods by exhibiting equivalent results.

Whenever the conjunctive method of combination of opinions is used, formulas are given that will require special numerical procedures supplied by software. However, whether the computation requires a few multiplies or a few thousand multiplies will have little impact on the actual computational time required for uncertainty reasoning since the main cost will be borne by the processes to compute constituent uncertainty values. The fact that sophisticated numerical methods might be required should not deter one from using the most appropriate combination formula possible. The computations that are required for conjunctive combination involve computing the mean and covariance of the minimum of multinormal random vectors. We can formulate these computations as follows.

Suppose that $\mathbf{x}^{(1)}$ is a random vector with mean $\boldsymbol{\mu}_1$ and covariance $\mathbf{C}_1$, and that $\mathbf{x}^{(2)}$ is a random vector with mean $\boldsymbol{\mu}_2$ and covariance $\mathbf{C}_2$. The corresponding density functions are given by

$$f_{\mathbf{x}^{(1)}}(\xi) = \frac{1}{(2\pi)^{N/2}[\det \mathbf{C}_1]^{1/2}} \exp\left[\frac{-(\xi-\boldsymbol{\mu}_1)^T \mathbf{C}_1^{-1}(\xi-\boldsymbol{\mu}_1)}{2}\right],$$

$$f_{\mathbf{x}^{(2)}}(\xi) = \frac{1}{(2\pi)^{N/2}[\det \mathbf{C}_2]^{1/2}} \exp\left[\frac{-(\xi-\boldsymbol{\mu}_2)^T \mathbf{C}_2^{-1}(\xi-\boldsymbol{\mu}_2)}{2}\right].$$

Now let $\mathbf{x} = \min(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ be a random vector defined by

$$\mathbf{x} = (\ \min(x(1), y(1))\ , \cdots , \ \min(x(N), y(N))\ ),$$

i.e., a random vector whose components are the minimum of the two respective random variables. Assuming that the sample space of $\mathbf{x}$ is obtained from independent trials from E_1 and E_2, then the product method results for the combining of experts. The mean and covariance of $\mathbf{x}$ can be computed from

$$\boldsymbol{\mu} = \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \min(\xi_1,\xi_2) f_{\mathbf{x}^{(1)}}(\xi_1) f_{\mathbf{x}^{(2)}}(\xi_2)\, d\xi_1\, d\xi_2, \tag{5.1}$$

$$\mathbf{C} = \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} [\min(\xi_1,\xi_2) - \boldsymbol{\mu}]^2 f_{\mathbf{x}^{(1)}}(\xi_1) f_{\mathbf{x}^{(2)}}(\xi_2)\, d\xi_1\, d\xi_2. \tag{5.2}$$

Although there are no trivial closed-form formulas, the computations are not as unformidable as it might first seem. We are aided by the fact that the marginal densities of $f_{\mathbf{x}^{(1)}}$ and $f_{\mathbf{x}^{(2)}}$ are Gaussian, and by the existence of fast numerical approximations to the error function (e.g., **erf**, which is a built-in function in some programming languages), defined by

$$\mathrm{erf}(x) = (2/\sqrt{\pi})\int_0^x \exp(-t^2)\,dt.$$

In the absolute worse case, numerical quadrature methods must be employed in order to find the resulting mean and covariance. Alternatively, an indeterminate approximation method might be invoked by generating a random collection of samples, and computing the resulting mean and covariances after taking pairwise minimums. In the tables, we denote the result of the multi-variate computation as

$$(\boldsymbol{\mu}, \mathbf{C}) = \phi(\boldsymbol{\mu}_1, \mathbf{C}_1, \boldsymbol{\mu}_2, \mathbf{C}_2),$$

where $\boldsymbol{\mu}$ is given by Equation (5.1), and $\mathbf{C}$ is given by (5.2).

When the combination method of the two bodies of experts is the either of the "combination by means" methods, then the updating involves a gaussian distribution with a constant value, rather than two gaussian distributions. The same formulas may be used, however, by using a zero covariance matrix for the constant distribution:

$$(\boldsymbol{\mu}, C) = \phi(\boldsymbol{\mu}_1, 0, \boldsymbol{\mu}_2, \mathbf{C}_2), \tag{5.3}$$

We will denote the resulting pair with the notation $(\boldsymbol{\mu}, \mathbf{C}) = \psi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \mathbf{C}_2)$.

Many of the formulas presented in the next three subsections involve considerable algebra for their derivations. We omit all such details, in the interest of brevity. In some cases, our terseness masks nonobvious manipulations.

5.1. Probabilistic opinions

We suppose that the opinions $x_\omega(\lambda)$ are estimates of the conditional probabilities $\mathrm{Prob}(\lambda|s)$, where s is the information known by the set of experts E. The means and covariances are then obtained from the $x_\omega(\lambda)$ data. The individual estimate vectors are $\mathbf{x}_\omega$, and are not necessarily probability vectors, i.e., the components may not sum to one. The mean opinion vector $\boldsymbol{\mu}$ contains the components $\mu(\lambda)$, and will likewise not necessarily be a probability vector. Of course, we can at any time normalize the individual opinions, or normalize the mean opinion. Although normalizing the mean opinion is not the same as taking the mean of the normalized opinions, it is not an unreasonable approximation, providing the opinions mostly lie close to the probability space. This is because the normalization process moves vectors in a direction that is approximately normal to the probability subspace. Similarly, the covariance matrix is not disturbed too badly providing the vectors are close to probability vectors.

If we assume conditional independence and unconditional independence, then the two estimates $\mathbf{x}^{(1)}_{\omega_1}$ and $\mathbf{x}^{(2)}_{\omega_2}$ combine using the formula $\mathbf{x}^{(1)}_{\omega_1}.\mathbf{x}^{(2)}_{\omega_2}.\boldsymbol{\gamma}^{-1}$, just as though all constituent estimates are probabilities. That is, the combined estimate for label λ is given by

$$x_{(\omega_1,\omega_2)}(\lambda) = \frac{\mathbf{x}^{(1)}_{\omega_1}(\lambda)\cdot\mathbf{x}^{(2)}_{\omega_2}(\lambda)}{\gamma_\lambda},$$

where γ_λ is the prior probability on λ (and is assumed to be a known constant). Providing the input vectors are approximately probability vectors, then the output will also be approximately a probability vector, because of the

Combining of Experts	Combining of Opinions		
	Conditional & Unconditional independence	α-dependence–	Conjunctive
Union	$\boldsymbol{\mu} = m_1\boldsymbol{\mu}_1 + m_2\boldsymbol{\mu}_2$ $\mathbf{C} = m_1\mathbf{C}_1 + m_2\mathbf{C}_2 + m_1(\boldsymbol{\mu}_1^2 - \boldsymbol{\mu}^2) + m_2(\boldsymbol{\mu}_2^2 - \boldsymbol{\mu}^2)$		
Product	$\boldsymbol{\mu} = \boldsymbol{\mu}_1.\boldsymbol{\mu}_2.\boldsymbol{\gamma}^{-1}$ $\mathbf{C} = (\boldsymbol{\gamma}^{-1})^2[\mathbf{C}_1\mathbf{C}_2 + \boldsymbol{\mu}_1^2\mathbf{C}_2 + \boldsymbol{\mu}_2^2\mathbf{C}_1]$	*	$(\boldsymbol{\mu},\mathbf{C}) =$ $\phi(\boldsymbol{\mu}_1,\mathbf{C}_1,\boldsymbol{\mu}_2,\mathbf{C}_2)$
Product measure	$\boldsymbol{\mu} = \boldsymbol{\mu}_1.\boldsymbol{\mu}_2.\boldsymbol{\gamma}^{-1}$ $\mathbf{C} = (\boldsymbol{\gamma}^{-1})^2[\mathbf{C}_1\mathbf{C}_2 + \boldsymbol{\mu}_1^2\mathbf{C}_2 + \boldsymbol{\mu}_2^2\mathbf{C}_1]$	*	$(\boldsymbol{\mu},\mathbf{C}) =$ $\phi(\boldsymbol{\mu}_1,\mathbf{C}_1,\boldsymbol{\mu}_2,\mathbf{C}_2)$
Pairwise matching	$\boldsymbol{\mu} = \boldsymbol{\mu}_1.\boldsymbol{\mu}_2.\boldsymbol{\gamma}^{-1}$ $\mathbf{C} = (\boldsymbol{\gamma}^{-1})^2[\mathbf{C}_1\mathbf{C}_2 + \boldsymbol{\mu}_1^2\mathbf{C}_2 + \boldsymbol{\mu}_2^2\mathbf{C}_1]$	*	$(\boldsymbol{\mu},\mathbf{C}) =$ $\phi(\boldsymbol{\mu}_1,\mathbf{C}_1,\boldsymbol{\mu}_2,\mathbf{C}_2)$
Noncomm. by means	$\boldsymbol{\mu} = \boldsymbol{\mu}_1.\boldsymbol{\mu}_2.\boldsymbol{\gamma}^{-1}$ $\mathbf{C} = (\boldsymbol{\gamma}^{-1})^2\boldsymbol{\mu}_1^2\mathbf{C}_2$	$\boldsymbol{\mu} = \boldsymbol{\mu}_1^{\alpha_2}.\boldsymbol{\mu}_2.\boldsymbol{\gamma}^{-\alpha_2}$ $\mathbf{C} = (\boldsymbol{\mu}_1^{\alpha_2}.\boldsymbol{\gamma}^{-\alpha_2})^2\mathbf{C}_2$	$(\boldsymbol{\mu},\mathbf{C}) =$ $\psi(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\mathbf{C}_2)$
Commutative by means	$\boldsymbol{\mu} = \boldsymbol{\mu}_1.\boldsymbol{\mu}_2.\boldsymbol{\gamma}^{-1}$ $\mathbf{C} = (\boldsymbol{\gamma}^{-1})^2[m_1\boldsymbol{\mu}_2^2\mathbf{C}_1 + m_1\boldsymbol{\mu}_1^2\mathbf{C}_2]$	$\boldsymbol{\mu} = m_1\boldsymbol{\mu}_1.\boldsymbol{\mu}_2^{\alpha_1}.\boldsymbol{\gamma}^{-\alpha_1}$ $+m_2\boldsymbol{\mu}_1^{\alpha_2}.\boldsymbol{\mu}_2.\boldsymbol{\gamma}^{-\alpha_2}$ $\mathbf{C} = m_1(\boldsymbol{\mu}_2^{\alpha_1}.\boldsymbol{\gamma}^{-\alpha_1})^2\mathbf{C}_1$ $+m_2(\boldsymbol{\mu}_1^{\alpha_2}.\boldsymbol{\gamma}^{-\alpha_2})^2\mathbf{C}_2$	$(\boldsymbol{\mu},\mathbf{C}) = m_1\psi(\boldsymbol{\mu}_2,\boldsymbol{\mu}_1,\mathbf{C}_1)$ $+m_2\psi(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\mathbf{C}_2)$

Table 3. Uncertainty calculi for probabilistic opinions. The entries marked '*' mean that a formula is not possible for this case. Notation is explained in Table 2. Especially, we reiterate that matrix juxtaposition, such as in C_1C_2 and $\boldsymbol{\mu}_1^2C_2$, denote termwise multiplication, rather than normal matrix multiplication.

assumption of unconditional independence, as discussed in Section 4.4.1. Accordingly, we may track the mean and covariance of the opinions, and normalize the mean opinion when we are finished, or at infrequent intermediate stages. The result will be similar to tracking all opinions, normalizing probabilities upon every pair of updates.

The pairwise matching method of updating leads to the same results as the product and product measure methods. For pairwise matching, the full independence assumptions are required of the opinions of the experts from the two bodies of experts, as discussed in Section 4.3.4.

When we assume α-dependence, then a combination formula is not possible for product, product measure, and pairwise combination of experts. (Formulas will be possible in the case of log-probability opinions). This is because the average of the α-th power of a collection of values is not related to the α-th power of the average. If the statistics were based on a geometric mean instead of the arithmetic mean, then formulas could be developed. When we say that a formula is not possible, what we mean is that for the given updating method, and for the statistics that are being tracked to describe the distribution of opinions of the bodies of experts being combined, there is insufficient information to deduce the desired statistics of the combined opinions.

For noncommutative updating by means, we assume that every expert in E_2 combines with the mean opinion of the experts in E_1 using an $\alpha(s_2, s_1)$ relationship, so that the combined opinion, indexed by experts in E_2, is

$$x_\omega(\lambda) = \frac{[\mu_1(\lambda)]^{\alpha_2} \cdot x_\omega^{(2)}(\lambda)}{\gamma_\lambda^{\alpha_2}}.$$

Note that the α-dependence is placed on the mean opinion of the experts in E_1, and not on the opinions in E_2, which is why the $\alpha_2 = \alpha(s_2, s_1)$ dependence is assumed. This choice is required in order to achieve an updating formula based on means and covariances; in fact, for log-probabilities, we will make the opposite choice. Once again, we require an additional unconditional α-dependence assumption, in order to be assured that the results are approximately probability vectors, so that we may normalize the mean value in approximating the mean of the normalized values.

For commutative updating by means, every expert opinion $\mathbf{x}_\omega^{(2)}$ for $\omega \in E_2$ combines with the mean opinion of E_1, namely μ_1, using an $\alpha(s_2, s_1)$ relationship, and every expert opinion of $\mathbf{x}_\omega^{(1)}$ for $\omega \in E_1$ combines with the mean opinion of E_2, namely μ_2, using an $\alpha(s_1, s_2)$ relationship. Once again, the appropriate α-value is always placed on the mean opinion from the other set of experts, rather than on the opinion. The opposite choice will be used for log-probability opinions. Also recall that, as discussed previously, it may be unrealistic to assume that both an

$\alpha(s_1, s_2)$ relationship and an $\alpha(s_2, s_1)$ relationship exist simultaneously.

In Table 3 we present a matrix of the formulations. Each set of formulas defines the method for obtaining the new mean and new covariance, respectively μ and $\mathbf{C}$, from the input means and covariances, μ_1, μ_2, $\mathbf{C}_1$, and $\mathbf{C}_2$.

5.2. Log-probabilities

The case of representation of the opinions by log-probabilities, with mean and covariance statistics and product set updating assuming conditional independence, is discussed in [22]. The same situation, but using α-dependence was introduced in [30]. Thus some of the entries in Table 4 can be found in these works.

Recall that the log-probabilities are defined as the quantities:

$$L(\lambda | s) = \log\left[\frac{\text{Prob}(\lambda | s)}{\text{Prob}(\lambda)}\right],$$

and that the $x_\omega(\lambda)$ are estimates of this quantity. The values can be either positive or negative, and are unbounded. Once again, $\mu(\lambda)$ represents the mean the opinions $x_\omega(\lambda)$, and can be positive or negative; $C(\lambda,\lambda')$ are the components of the covariance matrix of the values, and will generally define a matrix giving the ellipsoidal one-sigma distribution boundary of a multinormal (Gaussian) distribution fit to the set of opinions.

When the log-probability opinions are combined, the result is computed as though the estimates were true $L(\lambda | s)$ values. Thus for unconditional independence, we have that

$$x_{(\omega_1,\omega_2)}(\lambda) = x^{(1)}_{\omega_1}(\lambda) + x^{(2)}_{\omega_2}(\lambda).$$

In reality, we should subtract $\log K$ on the right hand side, where

$$K = \frac{[\text{Prob}(s_1, s_2)]}{\text{Prob}(s_1)\cdot[\text{Prob}(s_2)]}.$$

However, if we assume unconditional independence, then $\log K$ is zero. If we don't assume unconditional independence, then we can still omit the $\log K$ term, because log-probabilities are defined modulo an identical additive constant, and $\log K$ is constant over λ. This is a major advantage of the log-probability representation.

For α-dependence, our formulas assume only an $\alpha(s_1, s_2)$ relationship, except for commutative updating by means, where a two-way α-dependence is assumed. In the normal situation, we combine the opinions of the experts E_1 with the opinions of experts E_2 with the α-influence on the E_2-opinions, yielding

Combining of Experts	Combining of Opinions		
	Conditional independence	α-dependence	Conjunctive
Union	$\mu = m_1\mu_1 + m_2\mu_2$ $C = m_1 C_1 + m_2 C_2 + m_1(\mu_1^2 - \mu^2) + m_2(\mu_2^2 - \mu^2)$		
Product	$\mu = \mu_1 + \mu_2$ $C = C_1 + C_2$	$\mu = \mu_1 + \alpha\mu_2$ $C = C_1 + \alpha^2 C_2$	$(\mu, C) =$ $\phi(\mu_1, C_1, \mu_2, C_2)$
Product	$\mu = \mu_1 + \mu_2$ $C = C_1 + C_2$	$\mu = \mu_1 + \alpha\mu_2$ $C = C_1 + \alpha^2 C_2$	$(\mu, C) =$ $\phi(\mu_1, C_1, \mu_2, C_2)$
Product measure	$\mu = \mu_1 + \mu_2$ $C = C_1 + C_2$	$\mu = \mu_1 + \alpha\mu_2$ $C = C_1 + \alpha^2 C_2$	$(\mu, C) =$ $\phi(\mu_1, C_1, \mu_2, C_2)$
Pairwise matching	$\mu = \mu_1 + \mu_2$ $C = C_1 + C_2$	$\mu = \mu_1 + \alpha\mu_2$ $C = C_1 + \alpha^2 C_2$	$(\mu, C) =$ $\phi(\mu_1, C_1, \mu_2, C_2)$
Noncommutative by means	$\mu = \mu_1 + \mu_2$ $C = C_2$	$\mu = \alpha_2\mu_1 + \mu_2$ $C = \alpha^2 C_2$	$(\mu, C) =$ $\psi(\mu_1, \mu_2, C_2)$
Commutative by means	$\mu = \mu_1 + \mu_2$ $C = m_1 C_1 + m_2 C_2$	$\mu = m_1[\mu_1 + \alpha_1\mu_2] + m_2[\alpha_2\mu_1 + \mu_2]$ $C = m_1\alpha_2^2 C_1 + m_2\alpha_1^2 C_2 +$ $m_1(\alpha_2\mu_1 + \mu_2 - \mu)^2 + m_2(\mu_1 + \alpha_1\mu_2 - \mu)^2$	$(\mu, C) =$ $m_1\psi(\mu_2, \mu_1, C_1) +$ $m_2\psi(\mu_1, \mu_2, C_2)$

Table 4. Uncertainty calculi for log-probabilistic opinions using notation explained in Table 2.

$$x_{(\omega_1,\omega_2)}(\lambda) = x_{\omega_1}^{(1)}(\lambda) + \alpha x_{\omega_2}^{(2)}(\lambda).$$

Once again, the true formula should include a subtraction of $\log K$ on the right, where K is given by

$$K = \frac{[\mathrm{Prob}(s_1, s_2)]^\alpha}{\mathrm{Prob}(s_1)\cdot[\mathrm{Prob}(s_2)]^\alpha}.$$

However, if there is unconditional α-dependence, then this term is zero, and in any case the term is unnecessary because it is independent of λ, and thus constitutes an identical additive constant over the label set.

For pairwise matching, we no longer need complete independence of the sets of opinions $x_\omega^{(1)}(\lambda)$ and $x_\omega^{(2)}(\lambda')$. Instead, it suffices to have the orthogonality condition as given in Section 4.3.4.

For non-commutative updating by means, each expert in E_2 updates with the mean opinion from E_1 using an $\alpha_1 = \alpha(s_1, s_2)$ relationship, yielding:

$$x_{\omega_2}(\lambda) = \mu_1(\lambda) + \alpha x_{\omega_2}^{(2)}(\lambda).$$

Note that this is a different choice than made for the probability opinions. The alternative, placing the α-dependence term on the mean opinion, may be treated similarly, and leads to different formulas. For commutative updating by means, the experts in E_2 use an $\alpha(s_2, s_1)$ to combine with the mean opinion in E_1, whereas the experts in E_1 use an $\alpha(s_1, s_2)$ relationship to combine with the mean opinion in E_2. We may write this as:

$$x_\omega(\lambda) = \begin{cases} \alpha_2 x_\omega^{(1)}(\lambda) + \mu_2(\lambda) & \text{for } \omega \in E_1 \\ \mu_1(\lambda) + \alpha_1 x_\omega^{(2)}(\lambda) & \text{for } \omega \in E_2 \end{cases}$$

Table 4 presents the matrix of formulas.

5.3. Odds formulation

We now suppose that there are two labels, λ_1 and λ_2, denoting respectively the truth of a proposition H or its falsity, $\neg$H. In this case, each expert, using the information s, estimates the odds:

$$O(\mathrm{H}|s) = \frac{\mathrm{Prob}(\mathrm{H}|s)}{\mathrm{Prob}(\neg\,\mathrm{H}|s)}$$

and this single value is his opinion x_ω. The mean opinion is an average value μ, and the variance of the opinions is the single value C. Thus μ represents a nominal odds, and C is the inverse confidence that is placed in that value. Note that both μ and C are scalars. Throughout, γ stands for the prior odds on H, i.e., $O(\mathrm{H})$, and is assumed to be a known constant.

Otherwise, the situation is similar to the probabilistic case. There is no need for an unconditional independence assumption at any time, since the odds are never normalized, nor intended for normalization. The formulas are given in Table 5.

Although the two-label case may seem like a severe restriction, in fact most reasoning systems and inferencing networks operate in this domain. Indeed, since the labels in a multiple label-case can generally be formulated as the leaf nodes of a binary tree, the multi-label case can typically be handled by a number of two-label binary propositions, each representing a node of the tree. Accordingly, if we can simultaneously infer a collection of two-label problems, then we implicitly can do a multi-label case. However, this leaves open the question of representing uncertainty in a probabilistic labeling of multiple leaf labels given uncertain probabilistic labels of nodes in a binary tree.

6. Kalman filtering

In computer vision, robotics, and various other application domains, Kalman filtering is now often considered to be the preferred method for handling uncertainty reasoning [33-35]. Kalman theory, a well-developed branch of filtering and estimation in stochastic processes, is generally offered as an alternative approach to uncertainty reason-

Combining of Experts	Combining of Opinions		
	Conditional independence	α-dependence	Conjunctive
Union	$\mu = m_1\mu_1 + m_2\mu_2$ $C = m_1 C_1 + m_2 C_2 + m_1(\mu_1^2 - \mu^2) + m_2(\mu_2^2 - \mu^2)$		
Product	$\mu = \mu_1 \cdot \mu_2 / \gamma$ $C = [C_1 C_2 + \mu_1^2 C_1 + \mu_2^2 C_2]/\gamma^2$	*	$(\mu, C) =$ $\phi(\mu_1, C_1, \mu_2, C_2)$
Product measure	$\mu = \mu_1 \cdot \mu_2 / \gamma$ $C = [C_1 C_2 + \mu_1^2 C_1 + \mu_2^2 C_2]/\gamma^2$	*	$(\mu, C) =$ $\phi(\mu_1, C_1, \mu_2, C_2)$
Pairwise matching	$\mu = \mu_1 \cdot \mu_2 / \gamma$ $C = [C_1 C_2 + \mu_1^2 C_1 + \mu_2^2 C_2]/\gamma^2$	*	$(\mu, C) =$ $\phi(\mu_1, C_1, \mu_2, C_2)$
Noncommut. by means	$\mu = \mu_1 \mu_2 / \gamma$ $\mathbf{C} = \mu_1^2 C_2 / \gamma^2$	$\mu = \mu_1^{\alpha_2} \cdot \mu_2 / \gamma^{\alpha_2}$ $C = (\mu_1^{\alpha_2} / \gamma^{\alpha_2})^2 \cdot C_2$	$(\mu, C) =$ $\psi(\mu_1, \mu_2, C_2)$
Commutative by means	$\mu = \mu_1 \mu_2 / \gamma$ $C = [m_1 \mu_2^2 C_1 + m_1 \mu_1^2 C_2]/\gamma^2$	$\mu = m_1 \mu_1 \mu_2^{\alpha_1} / \gamma^{\alpha_1}$ $+ m_2 \mu_1^{\alpha_2} \mu_2 / \gamma^{\alpha_2}$ $C = m_1 (\mu_2^{\alpha_1} / \gamma^{\alpha_1})^2 C_1$ $+ m_2 (\mu_1^{\alpha_2} / \gamma^{\alpha_2})^2 C_2$	$(\mu, C) = m_1 \psi(\mu_2, \mu_1, C_1)$ $+ m_2 \psi(\mu_1, \mu_2, C_2)$

Table 5. Uncertainty calculi for opinions based on odds. In this table, all values are scalars, in that μ denotes an average value for the opinion of the odds, and C denotes its variance.

ing, and is not connected to calculi of the sort developed in the previous sections. Treatment of the general theory may be found in standard texts such as [36]. In this section, we will relate Kalman filtering in the static case to the ''opinions of experts'' formulation developed in this paper, thereby establishing relationships between the Kalman theory and, for example, Dempster/Shafer theory, and providing guidelines for applicability. (Shenoy has also noted the relationship of these two theories [37].)

There are many ways of deriving the Kalman filter. We will treat only the static case, and use the framework of multiple opinions of experts. The main differences from the calculi as developed in Section 5 lie in the choices for the representation of the opinions and the updating method.

In the static case (which forms a very special case of the Kalman machinery), there is a single value or single vector value $\mathbf{x}$ that must be estimated. Further, we will assume that the observations are identity projections of the state $\mathbf{x}$ with noise. Thus we are given a sequence of observations $\mathbf{x}_i$, for $i=1,2, \cdots n$, together with the covariances of the noise, C_i, $i=1,2, \cdots n$. In this especially simple setting, the Kalman filter says that the updated state, and thus the best estimate of the value $\mathbf{x}$, is given by:

$$\hat{\mathbf{x}} = \sum_{i=1}^{n} CC_i^{-1}\mathbf{x}_i,$$

$$C^{-1} = \sum_{i=1}^{n} C_i^{-1}.$$

Essentially, the estimated state is a weighted sum of the constituent estimates, with linear (matrix) weights related to the inverse of the respective covariance matrices, left normalized to sum to the identity matrix. The updating may be done sequentially, so that successive measurements may be combined with the current state to yield the same final outcome. For example, if the initial state is represented by the pair $(\mathbf{x}_1, C_1)$, and then new measurement $\mathbf{x}_2$ is presented with its covariance C_2, then the updated state is given by $(\hat{\mathbf{x}}, C)$, with

$$\hat{\mathbf{x}} = CC_1^{-1}\mathbf{x}_1 + CC_2^{-1}\mathbf{x}_2,$$

with

$$C = (C_1^{-1} + C_2^{-1})^{-1}.$$

If this process is repeated n times, then the outcome is as given above for the combination of n states. Because we are only treating the static case, the combination process is associative and commutative.

Clearly, this calculus resembles some of the calculi in the tables of Section 4. It is different from all of them, however. In order to interpret this calculus in the ''opinions of experts'' framework, we must impose different choices on the components of the formulation.

In particular, we no longer use probabilities nor log-probabilities nor odds in this formulation. Instead, the opinions represent estimates of some measurement or vector of measurements. This is the fundamental difference with the Kalman filter, and drives the application domains. Clearly, the statistics that are maintained are the means and covariances, and for the sake of argument, we will assume that the method for combining sets of experts is the product method. Because the values are no longer representative of probabilities, nor of values related to probabilities, the updating of pairs of opinions is not based on Bayesian analysis or probabilistic theory. Instead, pairs of opinions are updated using a ''weighted average approach.'' In particular, any pair of experts ω_1 and ω_2 combine their respective opinions $\mathbf{x}_{\omega_1}^{(1)}$ and $\mathbf{x}_{\omega_2}^{(2)}$ using a linear sum

$$\mathbf{x} = A\mathbf{x}_{\omega_1}^{(1)} + B\mathbf{x}_{\omega_2}^{(2)}$$

where A and B are matrices constrained to sum to the identity matrix:

$$A + B = I.$$

Here and in the following, matrices act on vectors and other matrices as in normal matrix products (and not termwise multiplications, as in Section 5).

What matrices A and B should the experts choose to combine their opinions? The answer depends on a certain amount of "common knowledge" among the experts. In the same way that the experts in the updating by means know about the mean value among the other experts, we will assume that each expert participating in the updating knows the means and covariances of both the set of experts to which the expert belongs and the set of experts to which the expert is combining. In particular, both ω_1 and ω_2 know C_1 and C_2 and the fact that all other pairs share the same information. In this case, the pair (ω_1, ω_2) can predict, in advance, that the outcome state will have covariance $AC_1A^T + BC_2B^T$. Since they also know in advance that $A+B = I$, they can determine an optimal A and B. In particular, they can determine that the variation in A by a symmetric matrix S leads to a first variation of

$$A\cdot(C_1+C_2)\cdot S - C_2\cdot S,$$

so that whether they want to minimize the trace, determinant, or any other positive multilinear measure of the resulting covariance, they are led to deduce that

$$A = C_2\cdot(C_1+C_2)^{-1}, \quad B = C_1\cdot(C_1+C_2)^{-1}.$$

With simple algebra, these equations are equivalent to:

$$A = C\cdot C_1^{-1}, \quad B = C\cdot C_2^{-2},$$

with $C = (C_1^{-1} + C_2^{-1})^{-1}$.

Since every pair of experts (ω_1, ω_2) will deduce the same pair of matrices A and B, the updated mean and covariance matrices, based on the product of experts $E_1 \times E_2$ of experts E_1 having mean μ_1 and covariance C_1 and experts E_2 having mean μ_2 and covariance C_2 will be

$$\mu = C\cdot C_1^{-1}\mu_1 + C\cdot C_2^{-1}\mu_2,$$
$$C = (C_1^{-1} + C_2^{-1})^{-1}.$$

These are precisely the updating equations as given by the Kalman filter for the static case.

There is an alternative logic whereby each committee (ω_1, ω_2) can deduce that their updated value should equal $CC_1^{-1}\mathbf{x}_{\omega_1}^{(1)} + CC_2^{-1}\mathbf{x}_{\omega_2}^{(2)}$. In this formulation, the experts know that they are trying to estimate a constant vector $\mathbf{x}$. The committee (ω_1, ω_2) knows that the realization of a gaussian random vector with mean $\mathbf{x}$ and covariance C_1 yields the value $\mathbf{x}_{\omega_1}^{(1)}$, and the realization of a different gaussian random vector, also having mean $\mathbf{x}$ but this time having covariance C_2, yields $\mathbf{x}_{\omega_2}^{(2)}$. Under the assumption that their best estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ is itself a Gaussian random vector, they can deduce that the probability density function of $\hat{\mathbf{x}}$ is a Gaussian with mean $C\cdot C_1^{-1}\mu_1 + C\cdot C_2^{-1}\mu_2$ and covariance $C = (C_1^{-1} + C_2^{-1})^{-1}$. Accordingly, the maximum likelihood guess for **x** by committee (ω_1, ω_2) is the same linear combination:

$$\mathbf{x}_{(\omega_1,\omega_2)} = C \cdot C_1^{-1} \boldsymbol{\mu}_1 + C \cdot C_2^{-1} \boldsymbol{\mu}_2.$$

Based on these updates over the entire product space of experts, the same uncertainty calculus results as before. Once again, we have assumed that all committees of experts know both covariance matrices C_1 and C_2.

Our treatment has assumed the product combination of sets of experts. As before, the product measure combination gives the same results. It is also possible to extend the treatment to obtain a similar result using the commutative by means method. Likewise, the pairwise matching formulation gives the same result, providing we assume orthogonality of the deviations of the experts from their means. This is the more usual formulation of stochastic processes, although it is not necessarily the most obvious formulation for every application. In particular, note that although all experts are attempting to estimate **x**, the mean of the individual estimates $\{\mathbf{x}_i\}_{i=1}^{\infty}$ is $\boldsymbol{\mu}$, which may not necessarily equal **x**. Nonetheless, it is the components of the errors $\{(\mathbf{x}_i - \boldsymbol{\mu})\}_{i=1}^{\infty}$ that we assume are orthogonal between sets of experts.

In summary, we see that the Kalman updating equations for the static case can be interpreted in terms of combining opinions of experts. The values that are represented are measurements, and not probabilities. The statistics are means and covariances, and the combining of sets is by means of the product, product measure, or pairwise matching method. Finally, the updating of pairs of opinions is based on common knowledge of the covariances matrices of the constituent sets of experts, and can be formulated either as an optimization on the resulting covariance of opinions, or can also be formulated as a maximum likelihood estimate based on conditional distribution functions.

We see that the domain of applicability of the static Kalman filter is much different than the calculi developed in Section 5. The principal point is that the updating is not based on Bayesian formulas or updating of probabilities; rather, the system is attempting to estimate a fixed collection of measurements, which in the formulation that we have given, are not changing. The updating, accordingly, is essentially averaging, with the weighting inversely proportional to the covariance of the respective estimates.

7. Discussion

There are at least three different broad classes of ways to express uncertainty: (1) Allow the probability for propositions to express the level of uncertainty; (2) Treat uncertainty as a range of opinions, and track that range as the opinions update; (3) Treat uncertainty as a separate entity that is affixed to each probability, and update uncertainties through predetermined functions that provide measures of uncertainty for each given mix of evidence (i.e., extrinsic uncertainty). In this paper, we have focused on the intermediate level, class (2), although we track uncertainty as a separate entity affixed to each probability (as in method (3)). Class (1) is well-known to be inadequate to distinguish between uncertainty and improbability, whereas (3), which might in some cases be necessary for realism, can become quite complicated if all the functional dependencies comprising the external uncertainty are accurately modeled.

We've established a set of different uncertainty calculi based upon choices in the formulation of the basic components of systems that track multiple opinions relating to probabilities. We were motivated by the Dempster/Shafer calculus, which tracks boolean opinions of ''possibilities'' by maintaining complete statistics on spaces of experts that update by taking product sets of experts and update opinions by intersecting possibilities. We identify four major choices: the statistics to be maintained, the values that the opinions represent, the method of combining sets of experts to form a new set of experts, and the method to update individual opinions. For statistics, we generally rely on maintaining means and covariances. We represent either probabilities, log-probabilities, or odds, and update pairs of probabilities using either independence assumptions or alpha-dependence. We also include conjunctive updating of pairs of probabilities as an example of functional updating. Finally, we have considered product sets for creating new sets of experts, as well as product measure, pairwise matching, non-commutative updating by means, and commutative updating by means.

The uncertainty calculi developed in this paper can be useful for Bayesian networks in two ways. First, we can represent and update information within the network using one of the calculi discussed earlier. For example, rather than maintaining probabilities at each node, we might maintain log-likelihoods. Or, independence relations can be replaced by alpha-independence relations, where alpha might vary throughout the network. Then, the nodes can include a representation of the uncertainty at the node, viewed conceptually as the statistics of multiple opinions about the state of the node.

Second, we could define a method for combining two equivalent networks. That is, given two states of a Bayesian network, we can combine the two by combining pairs of values at each node, and using the outcome as new information for the node. (Actually, the information impacts the node through the use of a "virtual" node.) This method for combining two states can also apply when the Bayesian network has been modified to use an uncertainty calculus for propogating information.

In examining the formulas, we see that the product, product measure, and pairwise matching forms of combinations of experts are equivalent, assuming that the necessary orthogonality conditions are placed on the experts in the pairwise updating case. The reason for treating all three cases is to point out this equivalence, thereby relating considerably different formulations. The noncommutative combination by means is presumably not very useful, because the level of uncertainty is always totally dependent on the level of uncertainty of the latest set of experts. The commutative updating by means, on the other hand, is potentially more interesting. In general, the assumption of independence of the evidence in the updating of individual probabilities results in the phenomenon that the increase in uncertainty when combining sets of opinions is independent of the separation of the means of the two sets. This is also true for the case of alpha-dependence when the product combination (or equivalent forms) are used. However, in the case of alpha-dependence and commutative updating by means, and in all cases when conjunctive updating is used, the degree of uncertainty will be affected by the separation in the means of the constituent bodies of opinions.

Overall, the formulas bear a considerable resemblance to one another, and to the Kalman updating procedure as well. In general, means mix in a diffusive pattern, and uncertainty levels represented by covariance matrices combine together. If a more elaborate or subtle combination of uncertainty is desired, then it is likely that the full approach of modeling extrinsic uncertainty (option (3) above) will be required.

Our theory manages to integrate a wide range of calculi, and yet presents separate formulations for each set of formulas, due to the choices that must be made in the components of the systems. As opposed to building a single parameterized algebraic system that subsumes all approaches, we believe that our "menu approach" is the more realistic and intuitive way to understand the relationships between the various candidates for uncertainty reasoning. Most especially, a knowledge of the underlying assumptions and components of a system will aid in the sensible use of the calculus in realistic applications.

Acknowledgements

This work derives from work that was supported by grant N00014-85-K-0077, Work Unit NR 4007006 from the Office of Naval Research, and NSF grant IRI-8703335. Robert Hummel was supported as a visiting professor by Vrije Universiteit, Amsterdam, The Netherlands, during the research period for this paper. Subsequent support under Air Force AFAL contract F33615-89-C-1087 reference 87-02-PMRE is gratefully acknowledged. Thanks are also due to INRIA-Rocquencourt Project Epidaure, where Robert Hummel was a visitor during the preparation of the manuscript. Larry Manevitz is a frequent visitor to NYU's Courant Institute of Mathematical Sciences.

References

[1] J. Gordon and E. H. Shortliffe, "The Dempster-Shafer theory of evidence," in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project,* B. G. Buchanan and E. H. Shortliffe (Eds.), Reading, Massachusetts: Addison-Wesley (1984).

[2] R. Duda, P. Hart, and N. Nilsson, "Subjective Bayesian methods for rule-based inference systems," *Proceedings of the National Computer Conference* **(AFIPS) 45**, pp. 1075-1082 (1976).

[3] R. Miller, H. Poole, and J. Myers, "INTERNIST-1: An experimental computer-bsed diagnostic consultant for general internal medicine," *New England Journal of Medicine* **307**, pp. 468-470 (1982).

[4] S. Andreassen, M. Woldbye, B. Falck, and S. Andersen, "MUNIN: A causal probabilistic network for interpretation of electromyographic findings," *Proceedings of the 10th International Conference on AI* **IJCAI-87**, pp. 366-372 (1987).

[5] J. Quinlan, "INFERNO: A cautious approach to uncertain inference," *The Computer Journal* **26**, pp. 255-269 (1983).

[6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,* Morgan Kaufman, San Mateo (1988).

[7] Steven L. Tanimoto, *The Elements of Artificial Intelligence,* W. H. Freeman and Company, New York (1995). (Second Edition.)

[8] R. Neapolitan, *Probabilistic Reasoning in Expert Systems,* John Wiley and Sons, New York (1990).

[9] L. Kanal and J. Lemmer, *Uncertainty in Artificial Intelligence,* Elsevier Science Publishers (North-Holland) (1986). *Uncertainty in Artificial Intelligence 2*, Lemmer and Kanal (Eds.), 1988, *Uncertainty in AI 3*, Kanal, Levitt, and Lemmer (Eds.), 1989, and Volumes 4 and 5 are also available. The proceedings for the 1991 and 1992 Workshop are available from *Morgan Kaufman, Inc.*.

[10] I. J. Good, "The measure of a non-measurable set," pp. 319-329 in *Logic, Methodology, and Philosophy of Science*, E. Nagel, P. Suppes, and A. Tarski (Eds.), Stanford University Press (1962).

[11] Henry E. Kyburg, Jr., "Bayesian and non-bayesian evidential updating," *Artificial Intelligence* **31**, pp. 271-293 (March, 1987).

[12] L. J. Savage, *The Foundations of Statistics,* John Wiley & Sons, New York (1954).

[13] D. Kahneman, P. Slovic, and A. Tversky, *Judgement under uncertainty: Heuristics and biases,* Cambridge University Press, Cambridge (1982).

[14] J. Pearl, "Fusion, propagation and structuring in belief networks," *Artificial Intelligence* **29**, pp. 241-288 (1986).

[15] G. Shafer, *A mathematical theory of evidence,* Princeton University Press, Princeton, N.J. (1976).

[16] L. Zadeh, "Fuzzy sets," *Information and Control* **8**, pp. 338-353 (1965).

[17] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics* **38**, pp. 325-339 (1967).

[18] P. M. Williams, ''On a new theory of epistemic probability,'' *British Journal for Philosophy of Science* **29**, pp. 375-387 (1978).

[19] J. Pearl, ''Reasoning with belief functions: An analysis of compatibility,'' *International Journal of Approximate Reasoning* **4**, pp. 363-389 (1990).

[20] J. Yen, ''Gertis: A Dempster-Shafer approach to diagnosing hierarchical hypotheses,'' *Communications of the ACM*, pp. 573-585 (1989).

[21] D. Spiegelhalter, A. P. Dawid , S. Lauritzen , and R. Cowell., ''Bayesian Analysis in Expert Systems,'' *Statistical Science* **8**, pp. 219-283 (1993).

[22] Robert A. Hummel and Michael S. Landy, ''A statistical viewpoint on the theory of evidence,'' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 235-247 (1988).

[23] J. Pearl, ''Rejoinder to 'Reasoning with belief functions: An analysis of compatibility',' ' *International Journal of Approximate Reasoning* **6**, pp. 425-443 (1992).

[24] S. Lauritzen and D. Spiegelhalter, ''Local computation with probabilities in graphical structures and their applications to expert systems,'' *Joural of the Royal Statistical Society B* **50** (2),(1988).

[25] C.-H. Tzeng, ''A mathematical formulation of uncertain information,'' *Annals of Mathematics and Artificial Intelligence* **4**, pp. 69-88 (1991).

[26] E. Charniak and D. McDermott, *Introduction to Artificial Intelligence,* Addison-Wesley, Reading, MA (1987).

[27] J. Gordon and E. H. Shortliffe, ''A method of managing evidential reasoning in a hierarchical hypothesis space,'' Stanford Computer Science Department Technical Report, (1984).

[28] G. Reynolds, D. Strahman, N. Lehrer, and K. Kitchen, ''Plausible reasoning and the theory of evidence,'' Computer and Info Science Dept, University of Massachusetts at Amherst Technical Report 86-11, (1986).

[29] P. Ruymgaart and T. Soong, *Mathematics of Kalman-Bucy filtering,* Springer-Verlag, New York (1985).

[30] Robert Hummel and Larry Manevitz, ''Combining bodies of dependent information,'' *Tenth International Joint Conference on Artificial Intelligence*, (August, 1987).

[31] D. Heckerman, ''Probabilistic interpretations for MYCIN certainty factors,'' pp. 127-135 in *Uncertainty in Artificial Intelligence*, L. Kanal and J. Lemmer (Eds.), Elsevier Science Publishers (North-Holland) (1986).

[32] A. Kandel, *Fuzzy Techniques in Pattern Recognition,* Wiley Press, New York (1982).

[33] N. Ayache and O. Faugeras, ''Building, registrating, and fusing noisy visual maps,'' *Proceedings of the First International Conference on Computer Vision,* pp. 73-82 (1987).

[34] O. Faugeras, F. Lustman, and G. Toscani, ''Motion and structure from motion from point and line matches,'' *Proceedings of the First International Conference on Computer Vision,* pp. 25-34 (1987).

[35] L. Matthies, R. Szeliski, and T. Kanade, ''Incremental estimation of depth maps from image sequences,'' *Proceedings of the Computer Vision and Pattern Recognition Conference*, (1988).

[36] P. S. Maybeck, *Stochastic Models, Estimation, and Control, Vols. 1 and 2,* Academic Press, New York (1979).

[37] P. Shenoy, ''A valuation-based language for expert systems,'' *International Journal of Approximate Reasoning* **3**, pp. 383-411 (1989).