The Emergence of Ordered Belief from Initial Ignorance

Paul Snow

P.O. Box 6134
Concord, NH 03303-6134
paulsnow@delphi.com

Abstract

Some simple assumptions about prior ignorance, and the idea that a sufficiently arresting contrast in the likelihoods of evidence will elicit belief that one proposition is at least as belief-worthy as another, lead to a partial ordering of propositions without the use of any kind of prior probability. The partial ordering is *not* a posterior probability distribution, but does share some intuitively pleasing properties of a probability, such as complementarity. Deciding the order (if any) between two disjunctions depends only on the highest likelihood disjunct in each, and so query handling in partitioned domains is efficient. In the event that an ordinary probability distribution is required for coherent decision making, one can be quickly calculated from the partial order.

Introduction

Ignorance is the unwillingness to order any of two or more sentences according to their belief-worthiness, unless one sentence implies the other. The unwillingness may be a matter of choice, as when a scientist wishes to interpret evidence about rival hypotheses without taking into account any personal views about the prior likeliness of the various rivals (Berger and Berry 1988). Other times, the unwillingness may be involuntary, as when there is no simply no basis for holding any opinion about the relative likeliness of the sentences in question.

However it arises, ignorance is not faithfully represented by any single probability distribution over the sentences. Whatever probabilities are assigned to the sentences, those probabilities are ordered with respect to one another, even though the sentences themselves generally are not. (For discussion of similar problems when representing ignorance in other uncertainty calculi, see Shenoy 1993.)

The approach developed in this paper avoids the assessment of a prior probability distribution under ignorance. Nevertheless, the emergence of ordered belief from prior ignorance retains a distinctly probabilistic flavor.

Notation and Assumptions about Ignorance

The notation

S >e> T

will denote the condition that the believer asserts that sentence *S* is, with a warrant satisfactory to the believer, at least as belief-worthy as sentence *T* in light of evidence *e*. If evidence *e* does not lead the believer to assert such an ordering of sentences *S* and *T*, then we write

S ?e? T

Note that this is distinct from asserting the *contrary* of S >e> T, which would be holding that *S* is less belief worthy than *T*. The condition of having no relevant evidence is indicated by the particle *nil*, as in

S ?nil? T

which expression denotes that there is no ordering between some sentences *S* and *T* in the absence of evidence.

We shall assume that the sentences of interest belong to a partitioned domain, which is defined as follows:

Definition. A *partitioned domain* is a set comprising:

(i) the always-true sentence, denoted **true**

(ii) the always false sentence, denoted **false**

(iii) two or more mutually exclusive sentences, called *atoms*

(iv) well-formed expressions involving atoms, **or**, and parentheses, called *simple disjunctions*

(v) well-formed expressions involving simple disjunctions, **true**, **false**, **or**, **not**, and parentheses

We shall also assume throughout that the atoms in the domain are collectively exhaustive, that is, exactly one of the atoms is true. This additional assumption places little epistemological burden on the believer (at worst, it means that one of the atoms is "none of the other atoms are true"), and has the convenient effect that every sentence in the domain has an equivalent simple disjunction.

Our first assumptions about ignorance, and the conquest of ignorance by evidence express the following ideas. If no evidence has yet been observed, and the question of relative belief-worthiness is not answerable on logical grounds, then there is no satisfactory warrant to order one sentence ahead of another. Even after evidence has been observed, the question may remain open. Once a commitment to an ordering is made, then other commitments may be inferred by conditional probability considerations, or by a fundamental belief-ordering consistency principle of the kind discussed by Sugeno (unpublished dissertation, cited in Prade 1985). The formal assumptions are:

A1. (Lack of explicit non-trivial prior orderings) For any sentences S and T,

S >nil> T implies that T implies S.

A2. (Lack of implicit non-trivial prior orderings) Values for conditional probabilities and orderings among them are neither known nor assumed if those values or orderings imply non-trivial constraints on the prior probabilities.

A3. (Consistency) For all evidence e, including *nil*, and any sentences S, S', T and T',

if S' implies S, then S >e> S';

if S' implies S and S' >e> T, then S >e> T;

if T' implies T and S >e> T, then S >e> T'.

A4. (Impartiality) If S >e> T, and S' and T' are sentences, and S is exclusive of T , then

if S' is exclusive of T and p(e | S') >= p(e | S), then S' >e> T, and

if S is exclusive of T' and p(e | T) >= p(e | T'), then S >e> T'.

A5. (Recovery from ignorance about atoms) For exclusive atoms s and t, and non-nil evidence e where p(e | s) > 0, a necessary and sufficient condition for s >e> t is that

f(e, s, t) >= q [A5.1]

where q is a real number chosen by the believer, and f(, ,) is a real-valued function chosen by the believer which is increasing in p(e | s) and decreasing in p(e | t), and such that a necessary condition for [A5.1] to hold is that p(e | s) is strictly greater than p(e | t), and such that p(e | t) = 0 is not a necessary condition for [A5.1] to hold.

A6. (Quasi-additivity) For any sentences S, T, and U where (S **and** U) and (T **and** U) are both false, and for all evidence e, including *nil*,

S **or** U >e> T **or** U if and only if S >e> T

An Inference Rule for Overcoming Ignorance In Simple Disjunctions

Assumptions A3 and A4 have a strong consequence when the propositions of interest belong to a partitioned domain. It is easy to show that if *D* is a simple disjunction, then the conditional p(e | D) is a convex combination of the p(e | d)'s, the conditionals for the evidence given each of the atoms within *D*. Thus,

$$P(e \mid D) =< \max_{d \text{ in } D} p(e \mid d) \qquad [1]$$

Theorem 1. Let S and T be simple disjunctions which are mutually exclusive, and let s and t be atoms where p(e | s) and p(e | t) are the greatest conditional probabilities for non-nil evidence e given atoms in S and T respectively.

S >e> T if and only if s >e> t.

Proof. S >e> T implies s >e> T by A4 and [1], which implies s >e> t by A3. Conversely, s >e> t implies s >e> T by A4 and [1], which implies S >e> T by A3. //

The theorem and assumption A5 lead to the following rule for deciding whether observed evidence *e* bearing on the states supports the assertion of S >e> T under certain circumstances:

Inference Rule. If S and T are simple disjunctions with no states in common, and if s and t are such that p(e | s) and p(e | t) are the greatest conditional probabilities for the evidence e given any atom in S and T respectively, then a sufficient condition for S >e> T is that f(e, s, t) >= q, where f(, ,) and q are as described in assumption A5.

This inference rule is strong enough by itself to handle problems like statistical hypothesis testing, where typically, disjoint propositions are compared, and often only one pair of propositions in a domain is of interest at any one time. Some further development to be introduced later will use the rule in a decision procedure which is applicable to all non-trivial ordering questions in partitioned domains.

Partial Qualitative Probability

Definition. A *partial qualitative probability* is a partial order of the sentences in a partitioned domain, such that, for all evidence e, including nil, and any sentences S, T, and U:

(i) (boundedness) **true** >e> S and S >e> **false**

(ii) (transitivity) (S >e> T) **and** (T >e> U) implies that S >e> U

(iii) (quasi-additivity) if S **and** U and T **and** U are both false, then

(S **or** U) >e> (T **or** U) if and only if S >e> T.

This definition is designed to echo that of an ordinary qualitative probability (de Finetti, 1937), differing only in being a partial, rather than a complete, ordering.

Within a partial qualitative probability, any ordering question involving simple disjunctions can be resolved by the theorem and the inference rule. To decide whether S >e> T:

(1) Eliminate from S and T all the atoms common to both, leaving S* and T*.

(2) If S* and T* are both empty, then S>e> T; if S* is empty and T* is not, then **not** (S >e> T); if T* is empty and S* is not, then S >e> T. Otherwise, apply the Inference Rule derived from Theorem 1 to S* and T*; S >e> T just in case S* >e> T*.

Partial qualitative probabilities also share an intuitively appealing property with ordinary probability distributions:

Theorem 2. (Complementarity) If S and T are simple disjunctions, and ">e>" is a partial qualitative probability, then

S >e> T implies **not** (T) >e> **not** (S)

Proof. Let C be the disjunction of atoms common to S and T, S' be the atoms in S and not in T, and T' be the atoms in T and not in S. Then by quasi-additivity, S' >e> T'. Let Q be the disjunction of atoms not in S and not in T. **not** (S) is T' **or** Q, and **not** (T) is S' **or** Q. Since S' >e> T', then by quasi-additivity, (S' **or** Q) >e> (T' **or** Q), or **not** (T) >e> **not** (S). //

Any Ordering Satisfying A1-A6 is a Partial Qualitative Probability

Lemma. If A, B, C, and D are simple or empty (containing no atoms except those that are false given the evidence) disjunctions, and there is no atom in common between A and B, nor any atom in common between C and D, then

A >e> B **and** C >e> D implies (A **or** C) >e> (B **or** D)

Proof. If (B **or** D) implies (A **or** C), then the required ordering holds. Suppose that is not the case. If B is empty or D is empty, then the lemma is trivial. If A is empty, then B is empty, and if C is empty, then D is empty [A3]. Suppose none of them are empty. For orderings to be asserted, evidence must be non-nil. Let a, b, c, and d be the atoms such that p(e | atom) is greatest among atoms in A, B, C, and D respectively.

WOLG, suppose that p(e | a) >= p(e | c). Let AC and BD disjoin the atoms that are peculiar to (A **or** C) and (B **or** D) respectively. By A5 and theorem 3, f (e, c, d) >= q, and since the function is increasing in p(e | second argument), f (e, a, d) >= q. Since f (e, a, b) >= q as well, then a >e> [the atom in (B **or** D) with the greatest p(e | atom)]. By A5, atoms a and d have different p(e | atom)'s, and so they must be distinct, and a must be distinct from all other atoms in D for the same reason; a is distinct from all atoms in B by hypothesis. So, a is in AC. BD is not empty, because we suppose no implication, so theorem 3 applies. The required ordering follows from A6. //

Note that the property proven in the lemma is generally **untrue** in conventional probabilistic reasoning systems. That it holds for systems satisfying A1-A6 is closely related to theorem 3 and the inference rule. In the absence of prior information or logical grounds to resolve the question, what matters in the comparison of sentences is the best-supported atom peculiar to each sentence. Thus, even though *A or C* may have atoms in common with *B or D*, this does not disrupt the ranking of their best-supported atoms (unless there are no atoms peculiar to each sentence, in which case, the order is logically determined).

Theorem 6. Any ordering satisfying assumptions A1-A6 is a partial qualitative probability.

Proof. **Boundedness:** Since **false** implies S, so S >e> **false** by A3, and since S implies **true**, **true** >e> S by A3.

Quasi-additivity: Assumption A6.

Transitivity: Let A be the disjunction of the atoms common to each of S, T, and U, B the atoms common to S and T alone, C those common to S and U alone, D those for T and U alone, and S*, T*, and U* those atoms unique to S, T, and U respectively. If e is nil, then T implies S and U implies T, so U implies S, and transitivity holds. Suppose, then, that e is not nil. Let a be the maximum conditional probability for e among the atoms of A, and b, c, d, s, t, and u be the corresponding quantities for B, C, D, S*, T*, and U*, respectively.

By quasi-additivity, we have S >e> T implies C **or** S* >e> D **or** T*. Similarly, T >e> U implies B **or** T* >e> C **or** U*. We wish to show that B **or** S* >e> D **or** U*. Since C **or** S* and D **or** T* have no atom in common, and nor do B **or** T* and C **or** U*, we apply the lemma to get

C **or** S* **or** B **or** T* >e> D **or** T* **or** C **or** U*

which by quasi-additivity simplifies to

B **or** S* >e> D **or** U*

as required. //

A Note on Assumption A5

In the assumption, we required that p(e | s) be strictly greater than p(e | t) in order for s >e> t to hold when p(e | s) is positive. We now present an example where if A5 called for a weak inequality, then the resulting ordering would fail to be a partial qualitative probability.

All letters are as in the transitivity portion of the proof of the theorem of the last section, and once again, we have S >e> T and T >e> U. An assignment of values for the atomic conditional probabilities consistent with this, and the modification of A5 to allow ordering assertions on weak inequalities, is:

$$d = .5, s = .4, u = .5, b = .4, t = .6, \text{and } c = .6$$

It is easy to confirm that under a weak inequality rule, C **or** S* >e> D **or** T*, the quasi-additive condition for S >e> T, and B **or** T* >e> C or U*, the condition for T >e> U. If the ordering is transitive, then S >e> U, and if it is quasi-additive, then B **or** S* >e> D **or** U*, so by theorem 3, it must be that either *b* or *s* is no smaller than both *d* and *u*. Neither is the case, since *b* is less than *d* or *u*, and so is *s*.

Adoption of Priors from Ordered Beliefs

A partial qualitative probability ordering possesses many of the intuitively appealing properties of a probability distribution. Nevertheless, it lacks the coherence of beliefs thought to be demanded in practical decision making problems, and provided by probability distributions (Lindley 1982), or in weaker form by set estimates (Kyburg and Pittarelli 1992).

Faced a similar conflict between the demands of modeling beliefs with Dempster-Shafer-style belief functions and the demands of coherence in action under risk, Smets (et al. 1991; Dubois et al. 1993) has proposed a two-tier system of belief representation, his "Transferable Belief Model". Up until action is required, beliefs are represented by the less-than-fully coherent D-S formalism (Smets' "credal" phase). Once action is called for, the original formalism is mapped onto a probability distribution, and that probability is used for decision making (Smets' "pignistic" phase). Once called into action, the probability distribution is also subject to revision in the face of further evidence using Bayesian methods.

At some point, therefore, the user of the ignorance representation may find it expedient to convert the orderings revealed by the evidence into an ordinary probability estimate, to use that estimate for decision making, and to apply further evidence to it using Bayes' theorem in the usual way.

Because of the restricted form of possible orderings consistent with theorem 1 and partial qualitative probability, it is quite tractable to use the asserted orderings to derive a useful "surrogate" probability distribution when the number of atoms in the domain is finite. It is generally impossible to have a truly agreeing single probability distribution, i.e., some distribution in which p(S | e) >= p(T | e) if and only if S >e> T. That's because any probability distribution is a complete ordering, rather than the partial ordering that arises from the assumptions. But it is easy to compute a probability distribution where for every S >e> T, the probabilities are ordered p(S | e) >= p(T | e).

The permissible orderings entail a single system of simultaneous linear constraints, each (apart from the total probability constraint) either of the form

$$p(s) >= c$$

(for atoms s where there is no distinct atom t such that s >e> t) where c is a non-negative constant which doesn't depend on the atom s, or else of the form

$$p(s) >= \Sigma p(s')$$

(for atoms s where there is one or more t such that s >e> t) where the summation is over all atoms s' such that s >e> s'. Since any atom s is ordered ahead of the disjunction of all the atoms s' such that s >e> s', the system has exactly one more non-redundant constraint than the number of atoms in the domain (the single total probability constraint is the extra constraint).

In order for the system to be consistent, that is, to have any solution, the constant c is bounded above by some positive quantity. It is easy to show that if c is chosen to equal that upper bound, then the system has a unique solution. The following algorithm computes the permissible upper bound on c and the associated unique solution to the system with effort that is linear in the number of atoms under discussion.

Algorithm for Computing Maximal *c* and Corresponding Solution

For N atoms, establish arrays:

Weight [1..N]	For each atom, the multiple of c that satisfies the order constraints
Runsum [1..N]	For atom indexed I, the sum of Weight [1] through Weight [I]
Prob [1..N]	The conditional probabilities for the evidence given each atom

and scalar quantities:

Index	As the name implies, an Index
Cutoff	An index, the least value where f (P [Index], P [Cutoff]) < q
Last	The value of Runsum [Cutoff - 1], or 1 if

```
        Cutoff = 1
BEGIN
1. Sort Prob [ ] in ascending order.
2. Initialize Cutoff = Last = Weight [ 1 ] = Runsum [ 1 ]
   = 1.
3. for Index = 2 .. N
     while f ( Prob [ Index ], Prob [ Cutoff ] ) >= q
         Last := Runsum [ Cutoff ]
         Cutoff := Cutoff + 1
     end while
     Weight [ Index ] := Last
      Runsum [ Index ] := Weight [ Index ]
                              + Runsum [ Index -1 ]
   end for
4. The maximum possible value of c is 1 / Runsum [ N ]; if
c is set to this maximum, then the unique solution of the
linear system is

      p( Index ) = Weight [ Index ] / Runsum [ N ].
END
```

Since *Cutoff* always increases in value, and never exceeds *Index*, it is easy to confirm that the effort required by the above algorithm is linear in the number of atoms.

Choosing Other Values for *c*

The single solution, maximum c approach is computationally simple, and places the least possible burden on subsequent evidence to overcome the low probability value assigned to the least favored atoms should one of them turn out to be true. On the other hand, smaller values of c may be preferred. In that case, the constraints describe a convex set of probability distributions, a set which contains all probability distributions which display all of the orderings asserted by the partial qualitative probability.

One reason for preferring a lower value of the constant c might be that the user prefers to use some particular other single probability distribution, for example, the maximum entropy distribution over all distributions consistent with the asserted ordering constraints. Such a distribution can be found using numerical or analytical optimization methods over the system with c = 0. Again, the simple form of the solution set, whether described in vertex or constraint form, should be an asset in searching for a congenial probability distribution. (There are exactly as many vertices as there are atoms, and the vertices are simple to enumerate using the information about Weight [] and Runsum [] produced by the algorithm of the last section.)

Another occasion for choosing a smaller c is when the user is content to represent beliefs for decision and action in convex set form. Although the convex set formalism lacks the full coherence of a singleton distribution, there is a considerable and growing literature which suggests methods for using convex sets in decision (see, for example, Sterling and Morrell 1991 for a review). Because of the small number of vertices, revision of the convex set in the light of further evidence is tractable (Levi 1980), and as with any convex set, revision can also be performed by a transformation of the system's coefficients (Snow 1991).

It can be shown that there are positive values of c such that the convex set represents *only* the orderings asserted by the partial qualitative probability. Among these, the largest such value will ordinarily be preferred since that choice places the least burden on subsequent evidence to reveal the truth of the least favored atoms should that happen to be necessary. Finding the largest such c requires about the same effort as enumerating the vertices with a known c, that is, order N^2. A full discussion of this point, however, is beyond the scope of the present paper.

Conclusions

Assumptions A1-A6 describe an intuitively appealing way that evidence can overcome initial ignorance. Although the mechanism is Bayesian, in that conditional probabilities are compared, there are no prior probabilities. Nevertheless, the inferences that arise from the assumptions retain some of the characteristics of probability distributions, including complementarity, and if normatively coherent behavior in gambling is required, then probabilities can be computed on demand. Query handling and the calculation of coherent probabilities are both computationally inexpensive.

References

Berger, J. O. and D. A. Berry, Statistical analysis and the illusion of objectivity, *American Scientist 76*, 159-165, 1988.

de Finetti, B., La prevision, ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincare 7*, 1-68, 1937 (English translation by H.E. Kyburg, Jr. in H.E. Kyburg, Jr. and H.E. Smokler, eds., *Studies in Subjective Probability*, New York: Wiley, 1964).

Dubois, D., H. Prade, and P. Smets, Representing partial ignorance, Workshop on Higher Order Uncertainty, George Mason University, July 1993.

Kyburg, H. E., Jr. and M. Pittarelli, Some problems for convex Bayesians, in D. Dubois, M.P. Wellman, B. D'Ambrosio, and P. Smets (eds.), *Uncertainty in Artificial Intelligence*, San Mateo, CA: Morgan Kaufmann, 149-154,1992.

Levi, I., *The Enterprise of Knowledge*, Cambridge, MA: MIT Press, 1980.

Lindley, D. V., Scoring rules and the inevitability of probability, *International Statistical Review 50*, 1-26 (with commentaries), 1982.

Prade, H., A computational approach to approximate and plausible reasoning with applications to expert systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence 7*, 260-283, 1985.

Shenoy, P.P., Modeling ignorance in uncertainty theories, Workshop on Higher Order Uncertainty, 1993.

Smets, P., Y-T. Hsia, A. Saffiotti, R. Kennes, H. Xu, and E. Umkehrer, The transferable belief model, in R. Kruse and P. Siegel (eds.), *Symbolic and Quantitative Approaches to Uncertainty*, Berlin: Springer-Verlag, Lecture Notes in Computer Science 548, 91-96, 1991.

Snow, P., Improved posterior probability estimates from prior and conditional linear constraint systems, *IEEE Transactions on Systems, Man, and Cybernetics 21*, 464-469, 1991.

Sterling. W. C. and D. R. Morrell, Convex Bayes decision theory, *IEEE Transactions on Systems, Man, and Cybernetics 21*, 173-183, 1991.