Probabilistic Temporal Reasoning

Thomas Dean* and **Keiji Kanazawa**
Department of Computer Science
Brown University
Box 1910, Providence, RI 02912

Abstract

Reasoning about change requires predicting how long a proposition, having become true, will continue to be so. Lacking perfect knowledge, an agent may be constrained to believe that a proposition persists indefinitely simply because there is no way for the agent to infer a contravening proposition with certainty. In this paper, we describe a theory of causal reasoning under uncertainty. Our theory uses easily obtainable statistical data to provide expectations concerning how long propositions are likely to persist in the absence of specific knowledge to the contrary. We consider a number of issues that arise in combining evidence, and describe an approach to computing probabilistic assessments of the sort licensed by our theory.

1 Introduction

The common-sense law of inertia [McCarthy, 1986] states that a proposition once made true remains so until something makes it false. Given perfect knowledge of initial conditions and a complete predictive model, the law of inertia is sufficient for accurately inferring the persistence of effects. In most circumstances, however, our predictive models and our knowledge of initial conditions are less than perfect. The law of inertia requires that, in order to infer that a proposition ceases to be true, we must predict an event with a contravening effect. Such predictions are often difficult to make. Consider the following examples:

- a cat is sleeping on the couch in your living room
- you leave your umbrella on the 8:15 commuter train
- a client on the telephone is asked to hold

In each case, there is some proposition initially observed to be true, and the task is to determine if it will be true at some later time. The cat may sleep undisturbed for an hour or more, but is extremely unlikely to remain in the same spot for more than six hours. Your umbrella will probably not be sitting on the seat when you catch the train the next morning. The client will probably hold for a few minutes, but only the most determined of clients will be on the line after 15 minutes. Sometimes we can make more accurate predictions (*e.g.*, a large barking dog runs into the living room), but, lacking specific evidence, we would like past experience to provide an estimate of how long certain propositions are likely to persist.

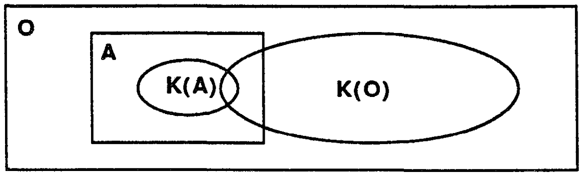

Figure 1: Precipitating events

Events precipitate change in the world, and it is our knowledge of events that enables us to make useful predictions about the future. For any proposition P that can hold in a situation, there are some number of general sorts of events (referred to as *event types*) that can affect P (*i.e.*, make P true or false). For any particular situation, there are some number of specific events (referred to as *event instances*) that occur. Let O correspond to the set of events that occur at time t, A correspond to that subset of O that affect P, $K(O)$ that subset of O known to occur at time t, and $K(A)$ that subset of A whose type is known to affect P. Figure 1 illustrates how these sets might relate to one another in a specific situation. In many cases, $K(O) \cap K(A)$ will be empty while A is not, and it may still be possible to provide a reasonable assessment of whether or not P is true at t. In this paper, we provide a probabilistic account of how such assessments can be made.

2 Prediction and Persistence

In the following, we distinguish between two kinds of propositions: propositions, traditionally referred to as *fluents*, which, if they become true, tend to persist without additional effort, and propositions, corresponding to the occurrence of events, which, if true at a point, tend to precipitate or trigger change in the world. Let $\langle P,t\rangle$ indicate that the fluent P is true at time t, and $\langle E,t\rangle$ indicate that an event of type E occurs at time t. We use the notation E_P to indicate an event corresponding to the fluent P becoming true.

Given our characterization of fluents as propositions that tend to persist, whether or not P is true at some time t may depend upon whether or not it was true at some $t-\Delta$, where $\Delta > 0$. We can represent this dependency as follows:

$$p(\langle P,t\rangle) = p(\langle P,t\rangle \mid \langle P,t-\Delta\rangle)p(\langle P,t-\Delta\rangle) + p(\langle P,t\rangle \mid \neg\langle P,t-\Delta\rangle)p(\neg\langle P,t-\Delta\rangle) \quad (1)$$

The conditional probability $p(\langle P,t\rangle \mid \langle P,t-\Delta\rangle)$ is referred to as a *survivor* function in classical queuing theory

*This work was supported in part by the National Science Foundation under grant IRI-8612644 and by an IBM faculty development award.

[Syski, 1979]. Survivor functions capture the tendency of propositions to become false as a consequence of events with contravening effects; one needn't be aware of a specific instance of an event with a contravening effect in order to predict that P will cease being true. As an example of a survivor function,

$$p(\langle P,t\rangle \mid \langle P,t-\Delta\rangle) = e^{-\lambda\Delta} \tag{2}$$

indicates that the probability that P persists drops off as a function of the time since P was last observed to be true at an exponential rate determined by λ. It is possible to efficiently construct an appropriate survivor function by tracking P over many instances of P becoming true [Dean and Kanazawa, 1987]. Referring back to Figure 1, survivor functions account for that subset of A corresponding to events that make P false, assuming that $K(A) = \{\}$.

If we have evidence concerning specific events known to affect P (*i.e.*, $K(A) \cap K(O) \neq \{\}$), (1) is inadequate. As an interesting special case of how to deal with events known to affect P, suppose that we know about all events that make P true (*i.e.*, we know $p(\langle E_P,t\rangle)$ for any value of t), and none of the events that make P false. In particular, suppose that P corresponds to John being at the airport, and E_P corresponds to the arrival of John's flight. We're interested in whether or not John will still be waiting at the airport when we arrive to pick him up. Let $e^{-\lambda\Delta}$ represent John's tendency to hang around airports, where λ is a measure of his impatience. If $f(t) = p(\langle E_P,t\rangle)$, then

$$p(\langle P,t\rangle) = \int_{-\infty}^{t} f(z)e^{-\lambda(t-z)}dz \tag{3}$$

A problem with (3) is that it fails to account for information concerning specific events known to make P false. Suppose, for instance, that $E_{\neg P}$ corresponds to Fred meeting John at the airport and giving him a ride to his hotel. If $g(t) = p(\langle E_{\neg P},t\rangle)$, then

$$p(\langle P,t\rangle) = \int_{-\infty}^{t} f(z)e^{-\lambda(t-z)}\left[1 - \int_{z}^{t} g(x)dx\right]dz \tag{4}$$

is a good approximation in certain cases. Figure 2 illustrates the sort of inference licensed by (4).

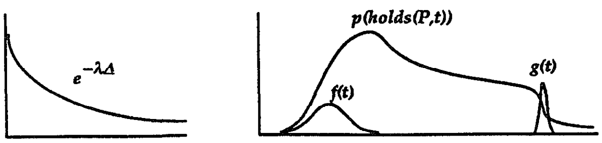

Figure 2: Probabilistic predictions

Equation (4) has problems also; in some cases, it counts certain events twice leading to significant errors. To combine the available evidence correctly, it will help if we distinguish the different sorts of knowledge that might be brought to bear on estimating whether or not P is true. The equation shown in Figure 3 makes the necessary distinctions and indicates how the evidence should be combined[1].

[1] In order to justify our use of the generalized addition law in (5), we assume that $p(\langle E_P,t\rangle \wedge \langle E_{\neg P},t\rangle) = 0$.

Consider the contribution of the individual terms corresponding to the conditional probabilities labeled N1 through N6 in (5). N1 accounts for *natural attrition*: the tendency for propositions to become false given no direct evidence of events known to affect P. N2 and N5 account for *causal accretion*: accumulating evidence for P due to events known to make P true. N2 and N5 are generally 1. N3 and N6, on the other hand, are generally 0, since evidence of $\neg P$ becoming true does little to convince us that P is true. Finally, N4 accounts for *spontaneous causation*: the tendency for propositions to suddenly become true with no direct evidence of events known to affect P.

$$\begin{aligned}
&p(\langle P,t\rangle) = && (5)\\
&\quad p(\langle P,t\rangle \mid \langle P,t-\Delta\rangle \wedge \neg(\langle E_P,t\rangle \vee \langle E_{\neg P},t\rangle)) && \text{(N1)}\\
&\qquad * \; p(\langle P,t-\Delta\rangle \wedge \neg(\langle E_P,t\rangle \vee \langle E_{\neg P},t\rangle)) &&\\
&\quad + p(\langle P,t\rangle \mid \langle P,t-\Delta\rangle \wedge \langle E_P,t\rangle) && \text{(N2)}\\
&\qquad * \; p(\langle P,t-\Delta\rangle \wedge \langle E_P,t\rangle) &&\\
&\quad + p(\langle P,t\rangle \mid \langle P,t-\Delta\rangle \wedge \langle E_{\neg P},t\rangle) && \text{(N3)}\\
&\qquad * \; p(\langle P,t-\Delta\rangle \wedge \langle E_{\neg P},t\rangle) &&\\
&\quad + p(\langle P,t\rangle \mid \neg\langle P,t-\Delta\rangle \wedge \neg(\langle E_P,t\rangle \vee \langle E_{\neg P},t\rangle)) && \text{(N4)}\\
&\qquad * \; p(\neg\langle P,t-\Delta\rangle \wedge \neg(\langle E_P,t\rangle \vee \langle E_{\neg P},t\rangle)) &&\\
&\quad + p(\langle P,t\rangle \mid \neg\langle P,t-\Delta\rangle \wedge \langle E_P,t\rangle) && \text{(N5)}\\
&\qquad * \; p(\neg\langle P,t-\Delta\rangle \wedge \langle E_P,t\rangle) &&\\
&\quad + p(\langle P,t\rangle \mid \neg\langle P,t-\Delta\rangle \wedge \langle E_{\neg P},t\rangle) && \text{(N6)}\\
&\qquad * \; p(\neg\langle P,t-\Delta\rangle \wedge \langle E_{\neg P},t\rangle) &&
\end{aligned}$$

Figure 3: Combining evidence about persistence

By using a discrete approximation of time and fixing Δ, it is possible both to acquire the necessary values for some of the terms N1 through N6 and to use them in making useful predictions [Dean and Kanazawa, 1987]. If time is represented as the integers, and $\Delta = 1$, we note that the law of inertia applies in those situations in which the terms N1, N2, and N5 are always 1 and the other terms are always 0. In the rest of this paper, we assume that time is discrete and linear and that the time separating any two consecutive time points is Δ. Only evidence concerning events *known* to make P true is brought to bear on $p(\langle E_P,t\rangle)$. If $p(\langle E_P,t\rangle)$ were used to summarize all evidence concerning events known to make P true, then N1 would be 1.

Before we consider the issues involved in making predictions using knowledge concerning N1 through N6, we need to add to our theory some means of predicting additional events. We consider the case of one event causing another event. The conditional probability

$$p(\langle E_2, t+\epsilon\rangle \mid \langle P_1 \wedge P_2 \ldots \wedge P_n, t\rangle \;\wedge\; \langle E_1,t\rangle) = \pi \tag{6}$$

indicates that, if an event of type E_1 occurs at time t, and P_1 through P_n are true at t, then an event of type E_2 will occur following t by some $\epsilon > 0$ with probability π. If the caused event is of a type E_P, this is often referred to as *persistence causation* [McDermott, 1982].

3 The Projection Problem

The *projection problem* [Dean and McDermott, 1987] involves computing the consequences of a set of conditions (observations) given a set of cause-and-effect relations referred to as *causal rules*. In [Dean and Kanazawa, 1987], we describe a probabilistic projection problem that naturally extends the deterministic version. The task in probabilistic projection is to assign each propositional variable of the form $\langle P, t\rangle$ a certainty measure consistent with the constraints specified in the problem. In this section, we provide examples drawn from a simple factory domain that illustrate the sort of inference required in probabilistic projection. We begin by introducing some new event types:

E_1 = "The mechanic on duty cleans up the shop"
E_2 = "Fred tries to assemble Widget17 in Room101"

and fluents:

P_1 = "The location of Wrench14 is Room101"
P_2 = "The location of Screwdriver31 is Room101"
P_3 = "Widget17 is completely assembled"

We assume that tools are occasionally displaced in a busy shop, and that P_1 and P_2 are both subject to an exponential persistence decay with a half life of one day; this determines N1 in equation (5). For $i = 1$ and $i = 2$:

$$p(\langle P_i, t\rangle \mid \langle P_i, t-\Delta\rangle \wedge \neg\langle E_{P_i}, t\rangle \wedge \neg\langle E_{\neg P_i}, t\rangle) = e^{-\lambda\Delta} \quad (7)$$

The other terms in (5), N2, N3, N4, N5, and N6, we will assume to be, respectively, 1, 0, 0, 1, and 0. When the mechanic on duty cleans up the shop, he is supposed to put all of the tools in their appropriate places. In particular, Wrench14 and Screwdriver31 are supposed to be returned to Room101. In the first example, henceforth Example 1, we assume that the mechanic is very diligent:

$$p(\langle E_{P_1}, t+\epsilon\rangle \wedge \langle E_{P_2}, t+\epsilon\rangle \mid \langle E_1, t\rangle) = 1.0 \quad (8)$$

Fred's competence in assembling widgets depends upon his tools being in the right place. In particular, if Screwdriver31 and Wrench14 are in Room101, then it is certain that Fred will successfully assemble Widget17.

$$p(\langle E_{P_3}, t+\epsilon\rangle \mid \langle P_1, t\rangle \wedge \langle P_2, t\rangle \wedge \langle E_2, t\rangle) = 1.0 \quad (9)$$

Let *T0* correspond to 12:00 PM 2/29/88, and *T1* correspond to 12:00 PM on the following day. Assume that ϵ is negligible given the events we are concerned with (*i.e.*, we will add or subtract ϵ in order to simplify the analysis).

$$\begin{aligned} p(\langle E_1, T0\rangle) &= 0.7 & (10)\\ p(\langle E_2, T1\rangle) &= 1.0 & (11)\end{aligned}$$

Let $BEL(A)$ denote an estimate of the likelihood of A given all of the evidence available. We are interested in assigning A a certainty measure consistent with the axioms of probability theory. We will sketch a method for deriving such a measure noting some, but not all, of the assumptions required to make the derivations follow from the problem specification and the axioms of probability. What can we say about $BEL(\langle P_3, T1+\epsilon\rangle)$? In this particular example, we can begin with

$$\begin{aligned} &BEL(\langle P_3, T1+\epsilon\rangle)\\ &= p(\langle E_{P_3}, T1+\epsilon\rangle)\\ &= p(\langle E_{P_3}, T1+\epsilon\rangle \mid \langle P_1 \wedge P_2, T1\rangle \wedge \langle E_2, T1\rangle)\\ &\quad * p(\langle P_1 \wedge P_2, T1\rangle \wedge \langle E_2, T1\rangle)\\ &= p(\langle P_1 \wedge P_2, T1\rangle \wedge \langle E_2, T1\rangle)\\ &= p(\langle P_1 \wedge P_2, T1\rangle)\end{aligned}$$

The last step depends on the assumption that the evidence supporting $\langle P_1 \wedge P_2, T1\rangle$ and $\langle E_2, T1\rangle$ are independent. The assumption is warranted in this case given that the particular instance of E_2 occurring at *T1* does not affect $P_1 \wedge P_2$ at *T1*, and the evidence for E_2 at *T1* is independent of any events prior to *T1*. Note that, if the evidence for E_2 at *T1* involved events prior to *T1*, then the analysis would be more involved. It is clear that $p(\langle P_1, T1\rangle) \geq 0.35$, and that $p(\langle P_2, T1\rangle) \geq 0.35$; unfortunately, we can't simply combine this information to obtain an estimate of $p(\langle P_1 \wedge P_2, T1\rangle)$, since the evidence supporting these two claims is dependent. We can, however, determine that

$$\begin{aligned} &p(\langle P_1 \wedge P_2, T1\rangle)\\ &= p(\langle P_1 \wedge P_2, T1\rangle \mid \langle P_1 \wedge P_2, T0\rangle) p(\langle P_1 \wedge P_2, T0\rangle)\\ &= p(\langle P_1 \wedge P_2, T0\rangle) * 0.5 * 0.5\\ &= p(\langle E_{P_1}, T0\rangle \wedge \langle E_{P_2}, T0\rangle) * 0.5 * 0.5\\ &= p(\langle E_{P_1}, T0+\epsilon\rangle \wedge \langle E_{P_2}, T0+\epsilon\rangle \mid \langle E_1, T0\rangle)\\ &\quad * p(\langle E_1, T0\rangle) * 0.5 * 0.5\\ &= 0.7 * 0.5 * 0.5\end{aligned}$$

assuming that there is no evidence concerning events that are known to affect either P_1 or P_2 in the interval from *T0* to *T1*.

To see how knowledge of events that provide evidence against certain propositions persisting is factored in, suppose that $T0 < T2 < T1$, and that there is a 0.1 chance that someone removed Wrench14 from Room101 at *T2* (*i.e.*, $p(\langle E_{\neg P_1}, T2\rangle) = 0.1$). In this example, henceforth Example 2, $p(\langle P_1, T1\rangle \mid \langle P_1, T0\rangle) = 0.5 * 0.9$, and, hence, $BEL(\langle P_3, T1+\epsilon\rangle) = 0.7 * 0.5 * 0.5 * 0.9$.

Another problem we have to address concerns events with consequences that are known to covary in a particular manner. In the next example, henceforth Example 3, we replace (8) in Example 1 with

$$\begin{aligned} p(\langle E_{P_1}, t+\epsilon\rangle \mid \langle E_1, t\rangle) &= 0.7 & (12)\\ p(\langle E_{P_2}, t+\epsilon\rangle \mid \langle E_1, t\rangle) &= 0.7 & (13)\end{aligned}$$

and (10) with $p(\langle E_1, T0\rangle) = 1.0$. Given the sort of analysis described above, we would calculate $BEL(\langle P_3, T1+\epsilon\rangle) = 0.35 * 0.35$ which is correct only assuming that the consequences of E_1 are independent. Suppose that we are explicitly told how the consequences of E_1 depend upon one another. For instance, if we are told in addition to (12) and (13) that

$$p(\langle E_{P_1}, t+\epsilon\rangle \wedge \langle E_{P_2}, t+\epsilon\rangle \mid \langle E_1, t\rangle) = 0.4$$

then an analysis similar to the one for Example 1 yields $BEL(\langle P_3, t+\epsilon\rangle) = 0.25 * 0.4$. In general, we assume that the consequences of any two events are independent unless we are explicitly told otherwise.

In the previous examples, there was at most one source of additional evidence that had to be considered at each step in combining all of the evidence concerning $\langle P_3, t+\epsilon\rangle$. In Example 4, we introduce two new fluents

$$\begin{aligned} P_4 &= \text{“Jack is on duty”}\\ P_5 &= \text{“Mary is on duty”}\end{aligned}$$

replace (8) in Example 1 with

$$\begin{aligned} p(\langle E_{P_1}, t+\epsilon\rangle \wedge \langle E_{P_2}, t+\epsilon\rangle \mid \langle P_4, t\rangle \wedge \langle E_1, t\rangle) &= 0.7\\ p(\langle E_{P_1}, t+\epsilon\rangle \wedge \langle E_{P_2}, t+\epsilon\rangle \mid \langle P_5, t\rangle \wedge \langle E_1, t\rangle) &= 0.9\end{aligned}$$

add that only one of Jack and Mary are ever on duty

$$p(\langle P_4, t\rangle \wedge \langle P_5, t\rangle) = 0.0$$

and provide information concerning who is likely to be on duty at *T0*

$$\begin{aligned} p(\langle P_4, T0\rangle) &= 0.4 \\ p(\langle P_5, T0\rangle) &= 0.3 \end{aligned}$$

In Example 4, we have $BEL(\langle P_3, t+\epsilon\rangle) = (0.3 * 0.45 * 0.45) + (0.4 * 0.35 * 0.35)$.

Throughout our analysis, we were forced to make assumptions of independence. In many cases, such assumptions are unwarranted or introduce inconsistencies. The inference process is further complicated by the fact that probabilistic constraints tend to propagate both forward and backward in time. This bi-directional flow of evidence can render the analysis described above useless..

4 Computing Projections

In [Dean and Kanazawa, 1987], we present a discrete approximation method for computing probabilistic projections using equations (3) and (4). Our method handles problems involving partially ordered events, but depends upon strong conditional independence assumptions. In this paper, we restrict our attention to totally ordered events, but consider a much wider class of constraints. We describe a method for computing a belief function for problems involving constraints of the form $p(A|B) = \pi$ corresponding to causal rules and the terms N1-6 in (5). Given that our method relies upon deriving a specific distribution, we begin by defining the underlying probability space.

Equations such as (7), (8), and (9) correspond to constraint schemata in which the temporal parameters are allowed to vary. For instance, in Example 1 we would have the following instance of (9)

$$p(\langle E_{P_3}, T1+\epsilon\rangle \mid \langle P_1, T1\rangle \wedge \langle P_2, T1\rangle \wedge \langle E_2, T1\rangle) = 1.0 \tag{14}$$

In the following, when we refer to the constraints in the problem specification, we mean to include all such instantiations of constraint schemata given all time points, plus additional marginals such as (10) and (11). The constraints in the problem specification define a set of propositional variables—infinite if the set of time points is isomorphic to the integers. We will assume that there are only a finite number of time points. For Example 1, *some* of the propositional variables are:

$x_1 \equiv \langle E_1, T0\rangle$ $\quad x_2 \equiv \langle E_{P_1}, T0+\epsilon\rangle$ $\quad x_3 \equiv \langle E_{P_2}, T0+\epsilon\rangle$
$x_4 \equiv \langle P_1, T0+\epsilon\rangle$ $\quad x_5 \equiv \langle P_2, T0+\epsilon\rangle$ $\quad x_6 \equiv \langle P_1, T1\rangle$
$x_7 \equiv \langle P_2, T1\rangle$ $\quad x_8 \equiv \langle E_2, T1\rangle$ $\quad x_9 \equiv \langle E_{P_3}, T1+\epsilon\rangle$

Using the complete set of propositional variables and constraints, we can construct an appropriate *causal dependency graph* [Pearl, 1986] that serves to indicate exactly how the variables depend on one another. The graph for Example 1 is illustrated in Figure 4 with the propositional variables indicated on a grid (*e.g.*, the variable $\langle P_1, T1\rangle$ corresponds to the intersection of the horizontal line labeled P_1 and the vertical line labeled *T1*).

Let $x_1, x_2, \ldots, x_n$ correspond to the propositional variables appearing in the causal dependency graph. We

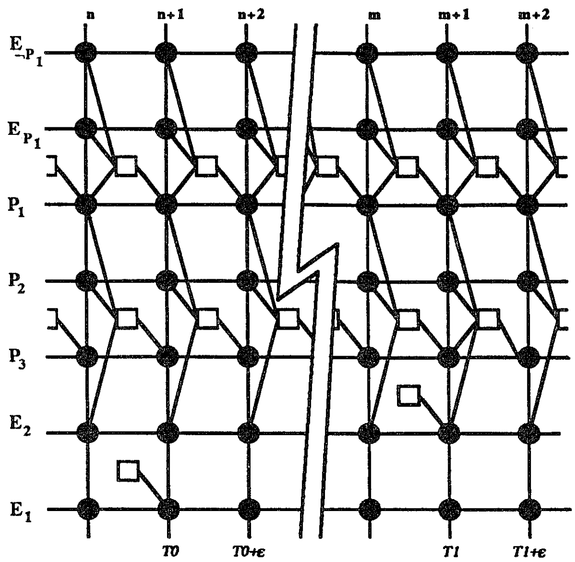

Figure 4: Causal dependency graph

allow a set C of boundary conditions corresponding to boolean equations involving the x_i's (*e.g.*, the constraint $\neg(\langle E_P, T1\rangle \wedge \langle E_{\neg P}, T1\rangle)$ might be represented as $(x_{14} = \mathsf{false} \vee x_{15} = \mathsf{false})$). The probability space Ω is defined as the set of all assignments to the x_i's consistent with C (*i.e.*, all $\omega = \langle x_1 = v_1, x_2 = v_2, \ldots, x_k = v_k\rangle$ such that $(v_i \in \{\mathsf{true}, \mathsf{false}\}) \wedge (\mathsf{consistent}(\omega, C))$). Each constraint specified in a problem can be expressed in terms of conditional probabilities involving the x_i's (*e.g.*, (14) might be encoded as $p(x_9 = \mathsf{true} \mid x_6 = \mathsf{true}, x_7 = \mathsf{true}, x_8 = \mathsf{true}) = 1.0$). Given

$$\begin{aligned} S_1 &= \{\omega \mid (\omega \in \Omega) \wedge (\mathsf{consistent}(\omega, A))\} \\ S_2 &= \{\omega \mid (\omega \in \Omega) \wedge (\mathsf{consistent}(\omega, B)\} \end{aligned}$$

we can rewrite $p(A|B) = \pi$ as $p(\omega \in S_1 \mid \omega \in S_2) = \pi$ or, using Bayes rule, we have

$$p(\omega \in S_1 \cap S_2) - \pi p(\omega \in S_2) = 0$$

from which we get

$$\sum_{\omega \in \Omega} (\mathcal{X}_{S_1 \cap S_2}(\omega) - \pi \mathcal{X}_{S_2}(\omega))\, p(\omega) = 0$$

where $\mathcal{X}_S$ is the indicator function for S (*i.e.*, $\mathcal{X}_S(\omega) = 1$ if $\omega \in S$, and 0 otherwise). Using the above transformations, we encode the problem constraints in terms of $a_1, a_2, \ldots, a_m$ where each a_i is a function of the form

$$a_i(\omega) = \mathcal{X}_{S_1^i \cap S_2^i}(\omega) - \pi_i \mathcal{X}_{S_2^i}(\omega)$$

where the S_j^i are derived following the example above. We now make use of the calculus of variations to derive a distribution p maximizing the entropy function

$$-\sum_{\omega \in \Omega} p(\omega) \log p(\omega)$$

over all distributions satisfying

$$\sum_{\omega \in \Omega} a_i(\omega) p(\omega) = 0 \qquad 1 \leq i \leq m$$

Using techniques described in [Lippman, 1986], we reduce the problem to finding the global minimum of the *partition function*, Z, defined as

$$Z(\lambda) = \sum_{\omega \in \Omega} \exp\left(-\sum_{i=1}^{m} \lambda_i a_i(\omega)\right)$$

where $\lambda = \langle \lambda_1, \ldots, \lambda_m \rangle$ and the λ_i's correspond to the Lagrange multipliers in the Lagrangian for finding the extrema of the entropy function subject to the conditions specified in the constraints[2]. Given certain reasonable restrictions on the a_i's[3], there exists exactly one $\bar{\lambda}$, corresponding to the global minimum of Z, at which Z has an extremal point. Given that Z is convex, we can use gradient-descent search to find $\bar{\lambda}$. Starting with some initial λ, gradient descent proceeds by moving small steps in the direction opposite the gradient as defined by

$$\nabla Z(\lambda) = \left| \frac{\partial Z}{\partial \lambda_1} \quad \frac{\partial Z}{\partial \lambda_2} \quad \cdots \quad \frac{\partial Z}{\partial \lambda_m} \right|$$

where

$$\frac{\partial Z}{\partial \lambda_k} = -\sum_{\omega \in \Omega} a_k(\omega) \exp\left(-\sum_{i=1}^{m} \lambda_i a_i(\omega)\right)$$

If $\|\nabla(Z)/Z\|$ goes below a certain threshold, the algorithm halts and the current λ is used to approximate $\bar{\lambda} = \langle \bar{\lambda}_1, \ldots, \bar{\lambda}_m \rangle$. We define a belief function BEL' as

$$BEL'(A) = \sum_{\omega \in \Omega_A} p(\omega)$$

where Ω_A is that subset of Ω satisfying A, and

$$p(\omega) = \frac{\exp\left(-\sum_{i=1}^{m} \bar{\lambda}_i a_i(\omega)\right)}{\sum_{\hat{\omega} \in \Omega} \exp\left(-\sum_{i=1}^{m} \bar{\lambda}_i(\hat{\omega})\right)} \qquad \forall \omega \in \Omega$$

Example	$BEL(\langle P_3, T1 \rangle)$	$BEL'(\langle P_3, T1 \rangle)$
1	0.1750	0.2058
2	0.1575	0.1908
3	0.1000	0.1078
4	0.1080	0.1417

Table 1: Comparing BEL and BEL'

Table 1 compares the results computed by BEL, the belief function discussed in Section 3, and the results computed by BEL'. Simplifying somewhat, BEL computes a certainty measure that corresponds to the greatest lower bound over all distributions consistent with the constraints, and, hence, $BEL(A) \leq BEL'(A)$. The certainty measure computed by BEL' is generally higher since the problems we are dealing with are underconstrained.

We should note that while the size of Ω is exponential in the number of propositions, Geman [Geman and Geman, 1984] claims to compute useful approximations using a method called *stochastic relaxation* that does not require quantifying over Ω. The results of Pearl [Pearl, 1986], Geman [Geman and Geman, 1984], and others seem to indicate that, in many real applications, there is sufficient structure available to support efficient inference. The structure imposed by time in causal reasoning presents an obvious candidate to exploit in applying stochastic techniques.

[2] Solving the Lagrangian directly, as in [Cheeseman, 1983], is made difficult by the fact that the equations obtained from the Lagrangian are nonlinear for constraints involving conditional probabilities.

[3] The most important restriction for our purposes being that the a_i's correspond to a set of linearly independent vectors.

5 Conclusions

We have presented a representational framework suited to temporal reasoning in situations involving incomplete knowledge. By expressing knowledge of cause-and-effect relations in terms of conditional probabilities, we were able to make appropriate judgements concerning the persistence of propositions. We have provided an inference procedure that handles a wide range of probabilistic constraints. Our procedure provides a basis to compare other methods, and also suggests stochastic inference techniques that might serve to compute useful approximations in practical applications. A more detailed analysis is provided in a longer version of this paper available upon request. In particular, we describe how observations that provide evidence concerning the occurrence of events and knowledge concerning prior expectations are incorporated into our framework.

References

[Cheeseman, 1983] Peter Cheeseman. A method of computing generalized bayesian probability values for expert systems. In *Proceedings IJCAI 8*. IJCAI, 1983.

[Dean and Kanazawa, 1987] Thomas Dean and Keiji Kanazawa. Persistence and probabilistic inference. Technical Report CS-87-23, Brown University Department of Computer Science, 1987.

[Dean and McDermott, 1987] Thomas Dean and Drew V. McDermott. Temporal data base management. *Artificial Intelligence*, 32:1–55, 1987.

[Geman and Geman, 1984] Stewart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[Lippman, 1986] Alan F. Lippman. *A Maximum Entropy Method for Expert System Construction.* PhD thesis, Brown University, 1986.

[McCarthy, 1986] John McCarthy. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[McDermott, 1982] Drew V. McDermott. A temporal logic for reasoning about processes and plans. *Cognitive Science*, 6:101–155, 1982.

[Pearl, 1986] Judea Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.

[Syski, 1979] Ryszard Syski. *Random Processes.* Marcel Dekker, New York, 1979.