Temporal Bayesian Networks

Ahmed Y. Tawfik and **Eric Neufeld**
Department of Computational Science, University of Saskatchewan
Saskatoon, Saskatchewan, Canada S7N 0W0

Abstract

Temporal formalisms are useful in several applications such as planning, scheduling and diagnosis. Probabilistic temporal reasoning emerged to deal with the uncertainties usually encountered in such applications. Bayesian networks provide a simple compact graphical representation of a probability distribution by exploiting conditional independencies. This paper presents a simple technique for representing time in Bayesian networks by expressing probabilities as functions of time. Probability transfer functions allow the formalism to deal with causal relations and dependencies between time points. Techniques to represent related time instants are distinct from those used to represent independent time instants but the probabilistic formalism is useful in both cases. The study of the cumulative effect of repeated events involves various models such as the competing risks model and the additive model. Dynamic Bayesian networks inference mechanisms are adequate for temporal probabilistic reasoning described in this work. Examples from medical diagnosis, circuit diagnosis and common sense reasoning help illustrate the use of these techniques.

1 Introduction

Bayesian networks are a probabilistic approach to reasoning about uncertainty. A Bayesian network [16] is a directed acyclic graph where nodes denote random variables and arcs represent causal dependencies between them. Throughout this paper we shall use a network, from [3], shown in Figure 1. An English equivalent for this network might be *'When the family goes out they turn on the light of the outdoor lamp and put the dog in the backyard. The dog's barking is heard when it is out in the backyard.'* This network has four random variables: *family-out, light-on, dog-out* and *hear-bark*. From the network structure, *family-out* affects the status of the light *light-on* and that of the dog *dog-out*. The arc between *dog-out* and *hear-bark* means that the dog's barking is heard when it is out. The topology of the graph represents the fact that the joint distribution of the variables can be written as the product of the conditional probability of each node given its immediate predecessors. From now on, fo, do, lo and hb stand for *family-out, dog-out, light-on* and *hear-bark* respectively. The joint probability distribution for the network in Figure 1 is $P(fo, lo, do, hb) = P(fo)P(lo|fo)P(do|fo)P(hb|do)$. The topology of the network together with the probability calculus allow the calculation of the probability of any random variable e.g. *family-out* given some evidence e.g. *light-on* and/or *hear-bark*. The probability of fo given lo is $P(fo|lo) = \frac{P(fo,lo)}{P(lo)}$ or $P(fo|lo) = \frac{P(fo,lo)}{P(lo,fo)+P(lo,\neg fo)}$ where $P(fo, lo) = P(lo|fo)P(fo)$ and $P(lo, \neg fo) = P(lo|\neg fo)P(\neg fo)$. Many AI and decision problems can be solved using Bayesian networks.

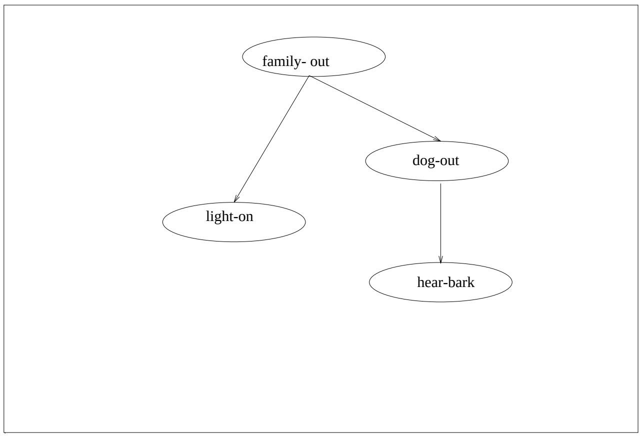

Figure 1: Bayes Network

Time is an important dimension for AI and reasoning. A first wave of investigators studied logics of time, see for example, Allen [1], McDermott [13] and McCarthy [12]. The study of probabilistic temporal formalisms is relatively new. Berzuini [2] uses Bayesian nets for temporal reasoning about causality. Berzuini's formalism considers each time period of interest as an additional random variable. This may considerably increase the number of variables in the network and the complexity of inference. Dean and Kanazawa [5] show how to represent persistence and causation. They use survivor functions to represent changing beliefs with time. This temporality allows predictive inferences but does not seem able to make inferences about independent time points. Dagum, Galper and Horvitz [4] use a dynamic belief network for forecasting. The dynamic network model shows how random variables at time t are affected on one hand by contemporaneous variables at time t as well as by the random variables at time $t-1$. Meixner [14] axiomatizes a probabilistic theory of possibility based on temporal propensity. Haddawy [7] introduces a probabilistic branching future model that corresponds to modal temporal logic.

2 Temporal Networks

A probabilistic temporal representation, like most other temporal representations, must address the issue of discrete versus continuous time. Discrete time is useful for various applications and seems simpler to deal with than continuous time. However interval based dense time seems more natural and elegant for reasoning. In temporal logics both representations are used , Allen [1] uses a dense time while McDermott [13] uses a point based time allowing intervals to be defined by its two end points. Probabilistic temporal representations have also used both time models, Berzuini [2] uses a continuous time and Dean and Kanazawa consider time to be discrete. Here, the position taken regarding this issue is to allow both, leaving the choice to the knowledge engineer. This is possible because discrete time corresponds to discrete probability and continuous time corresponds to continuous probability. Probability theory can handle both cases. In the continuous case, probability density functions are defined as functions of time. In the discrete case probabilities themselves are functions of time. For continuous time, the evaluation of the probability of the truth of certain fluents between two instants marking a period is the integral of the probability density function between those two instants. The discussion will deal mostly with discrete time but the equations for the continuous case will be mentioned and a circuit diagnosis example in Section 3.1 illustrates how to deal with probability density functions.

The second basic issue is how to associate probabilities and time. Two possibilities are considered: the first is to follow Berzuini's networks of dates model and consider times as random variables and the second is to parameterize probabilities with time. To motivate the discussion around this issue, let us introduce temporality to the *family-out* example in the previous section by adding the following statements:

1. If the family is out during the day they do not turn the light on,
2. if the family is not out during the day they open some windows.

Figure 2 illustrates the new network.

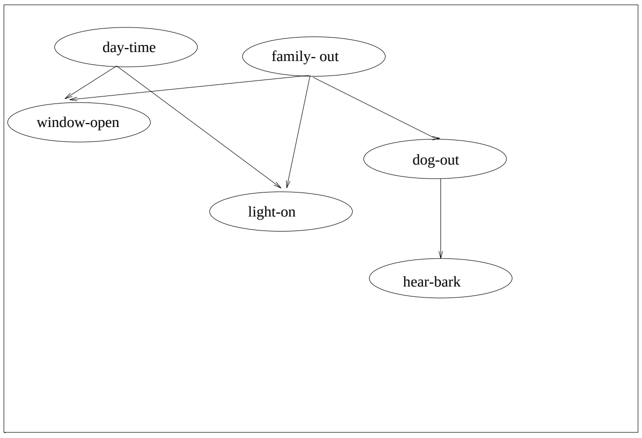

Figure 2: The Modified Network

The variable *window-open (wo)* represents the probability of the window being open. The temporal variable *day-time (dt)* is true when it is day and false otherwise. The joint probability distribution expressed by the network is $P(fo, dt, wo, lo, do, hb) = P(fo)P(dt)P(wo|fo, dt)P(lo|fo, dt)P(do|fo)P(hb|do)$. Representing day-time by a random variable complicates the network by increasing the number of nodes. The probabilities of wo and lo depend on the joint probability of dt and fo, this further complicates the representation. It may even get more cumbersome if we try to represent the following statements:

3. Usually the family is out between 9:00 am till 5:00 pm,
4. sometimes they come home for lunch between 12:00 noon and 1:00 pm,
5. when they come for lunch they do not bring the dog in,
6. they go to visit friends between 7:00 pm and 11:00 pm.

Moreover, temporal variables like dt do not seem to meet the definition of a random variable. Consider the rather intuitive definition of a random variable in [8]

> *'A random variable may be defined roughly as a variable that takes on different values because of chance.'*

Does *day-time* take on different values because of chance? The dt variable takes on the values 'true or false' depending on the deterministic motion of earth in space. Whether or not it is daytime can be determined from the sunrise and sunset times from the local newspaper. Treating time as a random variable complicates our reasoning unnecessarily. Here

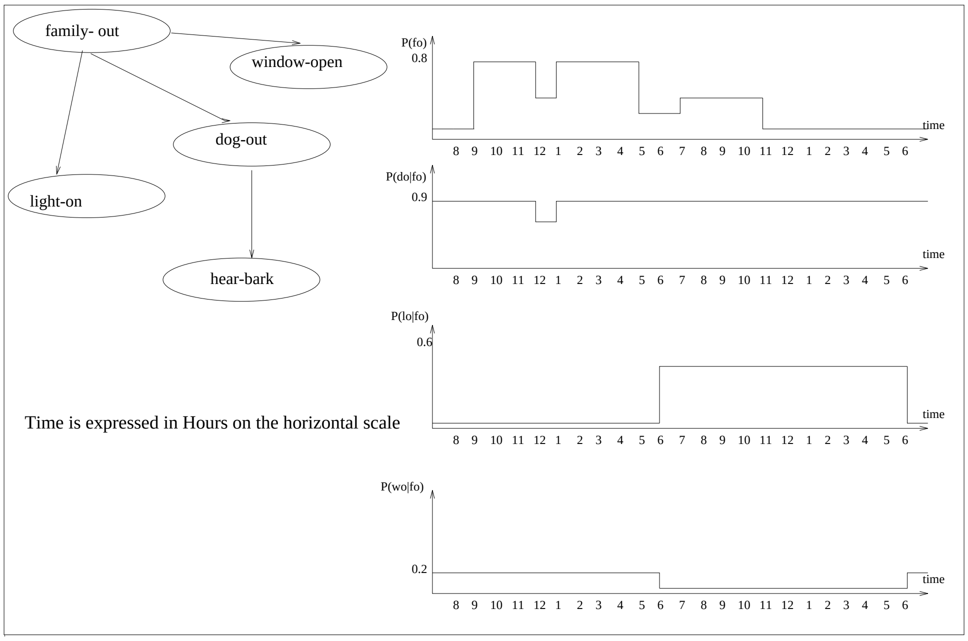

Figure 3: Probabilities as functions of time

the probabilities are functions of time, resulting in a simpler network.

Figure 3 illustrates this approach. The probabilities' variation with time is shown over a single day. The period 'a day' is a *complete cycle* after which probabilities follow the same pattern repeatedly. Such probability patterns capture the cyclic property of time useful in applications such as diagnosis of dynamic circuits with feedback. When the problem does not exhibit this cyclic property, the reasoning requires the expression of probability over a *window of interest*. In the next section, a circuit diagnosis example illustrates how to deal with acyclic time. Returning to Figure 3, $P(fo)$ captures conditions (3),(4) and (6) above. The change in $P(do|fo)$ between noon and 1:00 pm reflects statement (5). The first two statements are represented by the probabilities $P(lo|fo)$ and $P(wo|fo)$ where the daytime is from 6:00 am to 6:00 pm. The sharp changes in the probabilities in Figure 3 reflect the precision expressed in the statements. A smoother curve would be used to represent 'around 9 am' as opposed to '9:00 am'.

It is reasonable to assume $P(do|hb)$ is time independent. This may seem unusual here but note this representation does allow this flexibility. In Section 3.2 hb is treated as an event.

It is worth noting that although this formalism can answer questions of the form *'what is happening at time t?'* it cannot answer the question *'when does x happen ?'*. However in most applications, a direct probabilistic answer for the first question at different time points, approximates the probability distribution for the answer to the second, hence giving an indirect answer.

3 Temporal reasoning

From a temporal reasoning viewpoint, there are at least two types of probabilistic relationships between two different time points or periods; they may be completely independent of each other (unrelated), or dependent (related). If belief in fluent f at time t_i is not affected by the knowledge K at another time t_j then f at t_i is temporally independent of K at t_j. This independence means the probability of f at t_i and that of K at t_j are related by $P(f@t_i|K@t_j) = P(f@t_i)$.

In our example, if an observer goes past the house every few hours, instantaneously looks at the lights and checks if the dog is barking, then this observer can use the independence assumptions to reach a conclusion about *family-out*. Every observation is independent of the others provided that they are distant.

On the other hand, if the observer stays to watch the light and listen for the dog for few hours, then the reasoning in this case should be able to relate what happens at one instant with previous and future instants. In this case belief in fluent f at time t_i is affected by the knowledge K at time t_j and $P(f@t_i|K@t_j) \neq P(f@t_i)$. Persistence and causation are interesting special cases of reasoning about related time points.

3.1 Independent Time Instants or Periods

As mentioned before, the assumption of independence between different time points holds for the instantaneous observer. This means that observations made at time point t_i do not affect conclusions at t_j if $i \neq j$. If this observer sees the light and hears the dog barking then this observer can use the probabilities in Figure 3, to evaluate the probability of *family-out*. Later, new observations and conclusions are made that are completely independent of the previous ones because many events could have happened between the two time points.

For the instantaneous observer, temporal reasoning is therefore a set of atemporal Bayes nets except for the probabilities, they must correspond to the time point under consideration. Reasoning is not much harder than the atemporal case. To further illustrate the applicability of this assumption, consider the following circuit diagnosis problem.

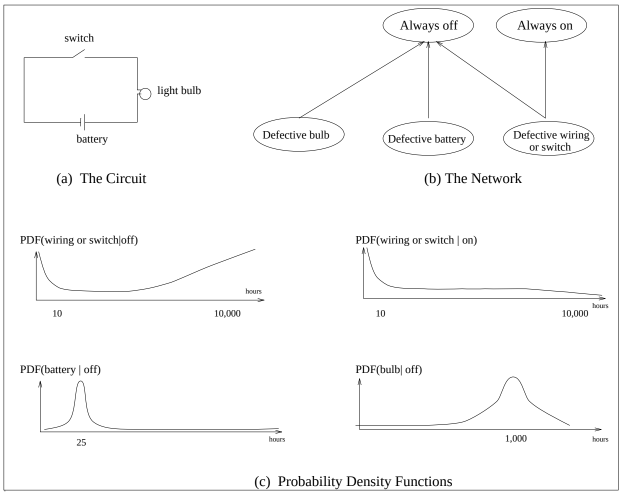

Figure 4: Circuit Diagnosis Example

> *A simple torch (flashlight) circuit consists of a bulb, a switch and a battery connected as in Figure 4-a. The probability distribution for the life time of the bulb and the battery are normal with means of 1000 and 25 hours respectively, as shown in Figure 4-c. The wiring and the switch rarely fail but their probability of failure is high initially due to burn-in faults. Then it drops as these defects usually affect the torch during the first few hours of operation. The failure probability finally rises again with aging.*

To evaluate the probability of *wiring-problem* given the torch is not working, it is necessary to know the number of operating hours after which the torch stopped working. The network used here, Figure 4-b, is similar to the symptom-disease network in [15]. Failures *defective switch/wiring, defective bulb* and *defective battery* cause or explain the malfunction of the torch. Two malfunctions can be observed: *always off* and *always on*. The window of interest here is the life time of the torch, expected to be some tens of thousands of hours. Replacement of a component can be represented by simply shifting the corresponding density function to start at the replacement time. Section 4 justifies such a shift within the temporal formalism. At time t, the probability malfunction M_j is caused by the failure of component c_i is given by

$$P(c_i|M_j) = \int_0^t PDF(c_i|M_j)dt$$

where PDF is the probability density function.

3.2 Dependent (Related) Time Instants or Periods

Now, suppose the observer, in the family-out example is monitoring the status of the light and the barking of the dog. If at time t_i the dog's barking is heard, one should conclude that the dog is out in the backyard at this instant and for some time afterwards. But after listening a while and not hearing the dog, one should be less certain about whether the dog is out or not. This decay in certainty with time is also a function of time that relates probabilities at all instants with an *event* where an *event* is an occurrence. In the present example any change in observations would be an event. Events tend to occur over time intervals of zero or longer duration. The distinction between event types and event tokens made by Hanks [9] is useful here. Event types are classes of events and event tokens are particular instances. While *hearing-bark* is an event type, *hearing-bark at 10:00* is an event token. Interaction between event tokens is discussed in Section 4.2.

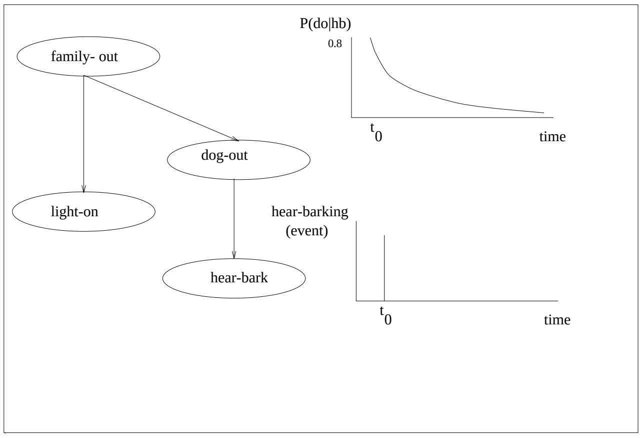

Figure 5: Probabilities dependence on events

For each event type, a probability transfer function represents the effect of an event token of this type on other variables. *A probability transfer function defines a relation between $P(x|e)$ and the time t for all t where x is a random variable, e an event type.*

The network in Figure 5 represents the decaying probability for *dog-out* given *hear-barking* using an exponential decay probability transfer function.

Changes affecting the same object may have different implications and hence different probability transfer functions. If the light is turned on somebody at home might have turned it on. Observing the light going off can have two possible explanations, the first that it was turned-off or it just burnt out. Observing the light going on is slightly stronger evidence that someone is at home than observing it going off. Figure 6 shows the network and the probabilities associated with such a *turn-light-on* scenario. The *light-turned-on* event is represented by a node that reduces the probability of *family-out*. Turning the light on instantaneously reduces belief in fo.

On the other hand this event token should affect *light-on* by making it true. The transfer function is a step function in this case. As time progresses, the observation that the light was turned-on some time ago does not contribute to the conclusion of *family-out* or otherwise. The arc marking this causal relationship is then either removed from the network or the conditional probability, as defined by the transfer function, saturates at a value chosen to mark indifference.

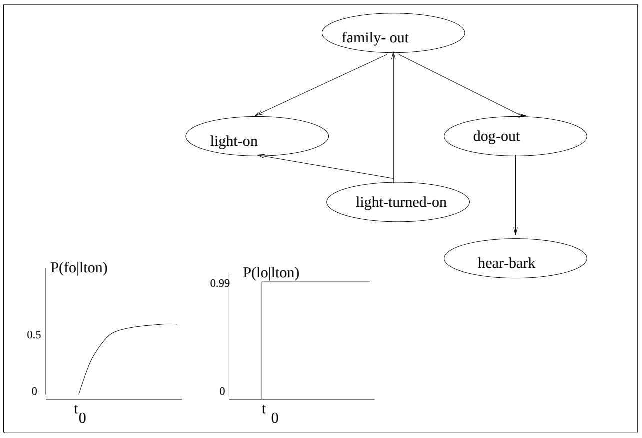

Figure 6: Light goes on

Probability transfer functions let us represent many forms of temporal dependencies. Consider the following example from [2] to see how this formalism expresses causal relations:

> *Either transplant or a subsequent transfusion may have caused an accidental inoculation of virus A or virus B. The inoculated virus, after a period of incubation, overgrows causing fever. Fever, however may also develop due to other causes.*

This can be represented as shown in Figure 7 and 8. The incubation period is represented as a delay between the infection and the fever and overgrowth. The probability of transfusion is high during and just after a transplant. This probability decreases exponentially with time as shown in Figure 8. The inoculation of virus A and B may happen any time during the transplant or the transfusion. The incubation period is different for virus A and virus B but in both cases has a normal distribution.

It is easy to calculate the probability $P(f \wedge o|virA)$ on a given day provided the time of transplant is known.

The distinction between monitoring and occasional sampling is a critical one. If occasional sampling is done frequently enough, it is equivalent to monitoring. The limit at which sampling can replace continuous monitoring, according to information theory, is equal to twice the maximum frequency in the signal. This may seem artificial in our present example but is useful in applications like diagnosis.

4 Convolutions, Probabilities and Models

It is necessary to unify the ideas temporal profile of random variables, conditional probabilities and the transfer function for events.

How should a transfer function combine with a temporal profile? Should hearing the dog barking continuously for five minutes, increase our certainty about the dog being out? What would be the certainty of *dog-out* if the dog just barked from time to time during the observation period?

Some mathematical tools and models are necessary to answer these questions.

4.1 Convolution in Probability Theory

The convolution integral of two distributions f_1 and f_2 is another distribution f written $f = f_1 \otimes f_2$. For continuous time, f is evaluated with the formula

$$f(t) = \int_{-\infty}^{\infty} f_1(t-\tau) f_2(\tau) d\tau.$$

or

$$f(n) = \sum_{m=0}^{\infty} f_1(m) f_2(n-m)$$

for discrete time.

If f_1 and f_2 represent the distribution of two random variables X and Y respectively, f is the distribution of a random variable $Z = X + Y$.

Now let f_1 be the distribution of a random variable t_1 defined as the time when we heard the dog bark, and let f_2 be the distribution of the random variable t_2 defined as the duration during which we continue to think that the dog is out after hearing the barking. Thus f allows us to evaluate the probability of *dog-out* following *hear-bark*. As shown in 4.2, convolution can also handle the cumulative effect of several events for additive storage models and a special case of the competing risks model.

Whenever the principle of superposition applies, that is when the effect of a set of inputs is the sum of the effect of each considered independently, the convolution can be used to evaluate the effect of a sequence of events.

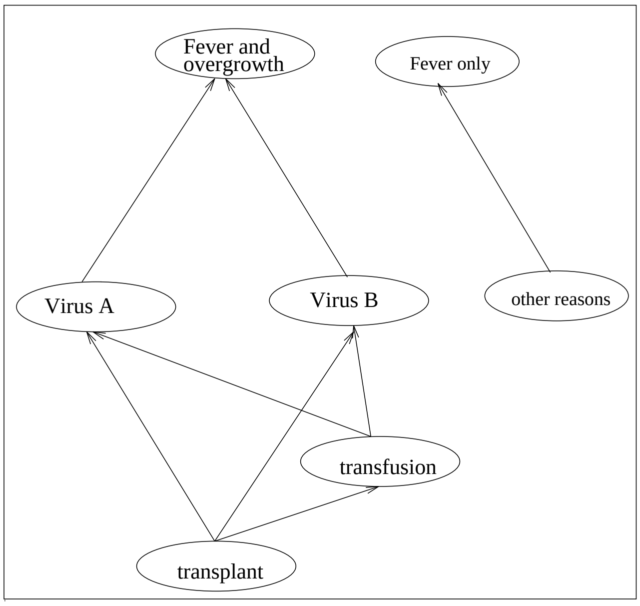

Figure 7: Medical Diagnosis example

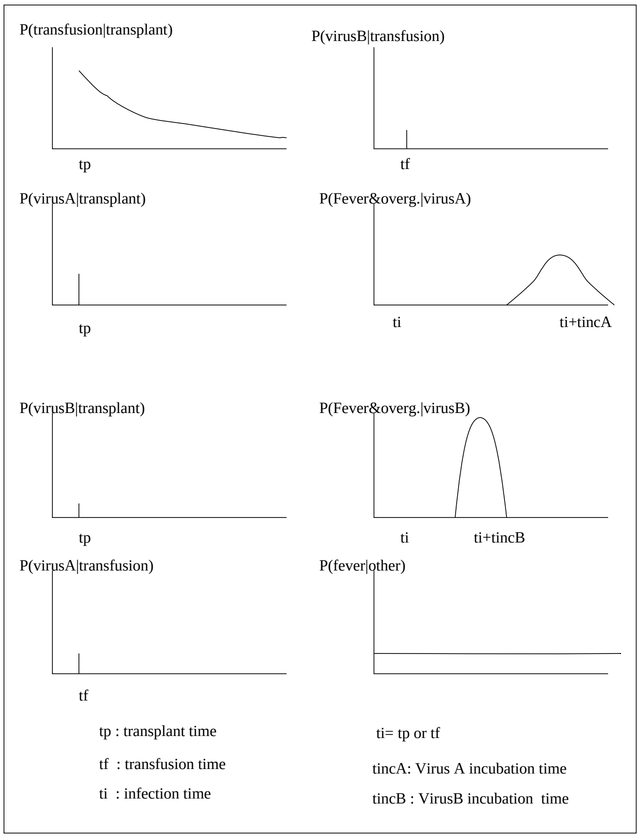

Figure 8: The Probabilities

4.2 Modeling Interaction of Events and Effects

Random variables representing beliefs, probability of events and effects of the events on belief interact in many interesting ways. In a Bayesian network representation, these interactions are reflected by dependencies in the joint probabilities. Consider for example the relation between *leaving-home, arriving-at-work* and *crowded-streets*. The probability of *arriving-at-work* at 9:00 as a result of *leaving-home* at 8:30 given that the streets are crowded is different from the probability if the the streets are not crowded. Possible causes for *crowded-streets* are *rain*, *accident* or *construction*. In general the transfer function can depend on time, *rain* causes *crowded-streets* only if it rains during the rush hour for more than 15 minutes, for example. There may be also causal dependencies between events: e.g. *rain* makes *accident* more probable. Dealing with this type of situation requires that the temporal profile of the joint probability distribution is known.

In some situations the interaction between events follows simpler models and the net effect of a number of events can be evaluated from the effect of a single event and information about the time at which they occurred. Storage processes, competing risks and domination are such models.

- **Storage process with additive inputs**

  A storage process can be thought of in terms of a warehouse or a dam, characterized by its inflow, its capacity and its release rule. See for example Glynn [6]. Let events be additive inputs, let release rules be functions in the inputs and let the storage level be the degree of belief such that the change in storage level reflects the change of belief with time. This model can represent our *dog-out* and *hear-bark* causal relation. Every token of *hear-bark* tends to fill the belief in *dog-out* to a certain level. The release rule guarantees an exponential decay of this belief. These systems follow the conservation of mass principle. This principle, when applied to our example implies that we cannot believe *dog-out* unless we hear barking at least once. It may be useful however to allow *do* to have a non zero probability before any *hb* event token. One way to do this is to use the $P(hb)$ as an input causing $P(do) > 0$.

  Convolution can be used to evaluate the belief (storage) resulting from the accumulation of events (inputs) of a storage process. Glynn [6] shows a storage process can be approximated by a finite state space model. Transfer functions can be derived from the state space model or designed to reflect the same behavior.

  Figure 9 illustrates how the probability of *dog-out* changes given different *hear-barking* event patterns, where probabilities are calculated us-

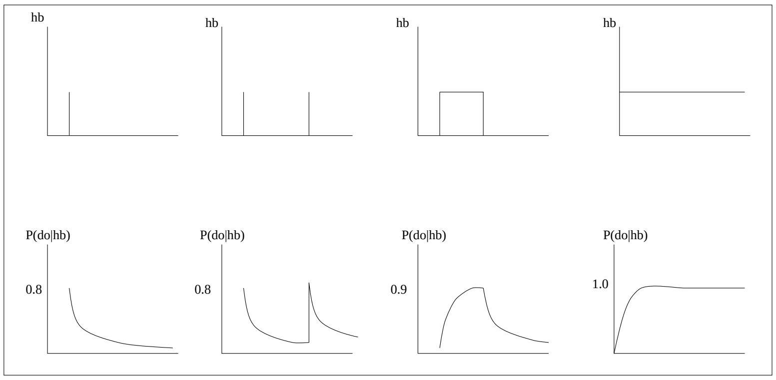

Figure 9: Convolution Results

ing convolution. In this figure, the belief in *dog-out* rises sharply whenever *hear-bark* takes place, and the degree of belief reached each time is slightly higher. If the barking is heard continuously over a period, the belief in *dog-out* keeps rising during this period and then decays after barking ceases to be heard. The fourth event pattern in the figure deals with the case when the dog is heard continuously. In this case belief rises and then almost saturates.

The use of two release rules allows us to compute the probability that the city on the horizon is Regina after driving from Saskatoon for time t. The first release rule lets *driving* accumulate until after you have arrived in Regina with high probability. Then the second rule is applied letting the probability of being in Regina decay with *driving* to express the idea that you should have passed Regina. Both rules may be non-linear such that the probability get a shape similar to the expected distribution.

- **Competing Risks Model**

  As the name suggests, the competing risks model represents two or more potential dangers competing to cause the failure of an organism. An interesting generalization of this model is when different potential causes compete to produce the same effect and the success of one of them prevents the others from successeding. Kalbfleisch and Prentice [11], and Hutchinson [10] consider some applications of this model. In the context of events, competitions occur frequently: e.g. repeated exposures to some viruses compete until one causes infection after which the body develops an immunity. For continuous time, the probability density of failure $f(t)$ due to two risk is $f(t) = f_0(t) + f_1(t)$. Here $f_0(t) = l_0(t) * L_1(t)$ and $f_1(t) = l_1(t) * L_0(t)$. $L_1(t)$ and $L_0(t)$ are the probabilities of survival at time t while $l_0(t)$ and $l_1(t)$ are the probability density of failure (hazard function). The case for discrete time is simpler and the probability is $f(n) = l_0(n) + l_1(n) - l_0(n) * l_1(n)$. $l_0(n)$ and $l_1(n)$ are the probability of failure at time n for the first and second risk respectively.

  In both cases if the effects of the two infections do not overlap in time we obtain a simpler form $f(t) = l_0(t) + l_1(t)$. If the two risks have the same hazard function $l(t)$ but at different times, which is the case if a person is exposed twice to the same virus then $l_0(t) = l(t-t_0), l_1(t) = l(t-t_1)$ where t_0 and t_1 are the times of exposure to the virus. In this case convolution can be used to evaluate the probabilities. Otherwise the use of the original equations would be required.

- **Dominating Events Model**

  In this model a particular event tends to dominate the others. Rules to determine the dominating event along with the transfer function of this event are needed in this context. The rules can be simple, like *the most recent event* and this rules applies to our circuit diagnosis example. Given a sequence of *changing-bulb* events, the probability that the circuit failed due to *burned-out-bulb* depends on the last *changing-bulb* and the lifetime of the bulbs. The condition for this model to apply is that the dominating event makes the dominated events irrelevant to the reasoning.

5 Conclusion

Representing probabilities as functions of time seems to be a simple and useful technique for implementing probabilistic temporal reasoning. Probability transfer functions can represent different types of events. Known Bayes net inference methods can evaluate probabilities of outcomes. By characterizing the dependencies between different instants or periods temporal reasoning can be done with a small computational overhead. Decoupling temporal reasoning across different time points from the reasoning at a given time point simplifies reasoning. Simple interaction models such as the storage model, competing risks model or the domination model can be used to represent some useful temporal phenomena without complicating inference.

6 Acknowledgements

The first author thanks the Institute for Robotics and Intelligent Systems (IRIS) and the University of Saskatchewan for support. Research of the second author is supported by IRIS and the Natural Science and Engineering Research Council of Canada (NSERC).

References

[1] J. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–154, July 1984.

[2] C. Berzuini. Representing time in causal probabilistic networks. In *Uncertainty in Artificial Intelligence 5*, pages 15–28, (North-Holland), 1990. Elsevier Science Publishers B.V.

[3] E. Charniak. Bayesian networks without tears. *AI Magazine*, 12(4):51–63, Winter 1991.

[4] P. Dagum, A. Galper, and E. Horvitz. Dynamic network models for forecasting. In *Proceedings of the 1992 Workshop on Uncertainty in Artificial Intelligence*, pages 41–48. Association for Uncertainty in Artificial Intelligence, 1992.

[5] T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5(3):142–150, August 1990.

[6] J. Glynn. A discrete-time storage process with a general release rule. *Journal of Applied Probability*, 26:566–583, 1989.

[7] P. Haddawy. A temporal probability logic for representing actions. In *Proc. of the 1991 Conference on Knowledge Representation*, pages 313–324, 1991.

[8] M. Hamburg. *Statistical Analysis for Decision Making,* 4^{th} *ed.* HBJ Publishers, Orlando, FL, 1987.

[9] S. Hanks. Representing and computing temporally scoped beliefs. In *Proc. of AAAI- 88, St. Paul, MN*, pages 501–505, 1988.

[10] T. Hutchinson. A note on applications of the competing risks model. *Accident Analysis and Prevention*, 15(3):225–226, 1983.

[11] J. Kalbfleisch and R. Prentice. *The Statistical Analysis of Failure Time Data.* John Wiley and Sons, New York, 1980.

[12] J. McCarthy and P. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*. Edinburgh University Press, Edinburgh, 1969.

[13] D. McDermott. A temporal logic for reasoning about processes and plans. *Cognitive Science*, 6(2):105–155, April 1982.

[14] U. Meixner. Propensity and possibility. *Erkenntnis*, 38(3):323, May 1993.

[15] J. Pearl. Distributed revision of composite beliefs. *Artificial Intelligence*, 33:173–215, 1988.

[16] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kauffmann, San Mateo, CA, 1988.