Towards Temporal Reasoning Using Qualitative Probabilities

Ahmed Y. Tawfik

School of Computer Science
University of Windsor
Windsor, Ontario N9B 3P4
Canada
atawfik@uwindsor.ca

Abstract

Previously proposed qualitative probabilities do not satisfy the requirements of temporal reasoning, and a new abstraction is proposed here that represents probability changes over time using four values: increasing, decreasing, constant, and unknown. The qualitative temporal probability presented satisfies the axioms of probability by obeying a set of probabilistic constraints. Temporal constraints are imposed through a set of causality, ordering and probabilistic constraints.

1. Introduction

Qualitative imprecise probabilities accomplish many reasoning tasks without exact numeric calculations. Instead, these qualitative probabilities impose constraints implied by the probability theory on symbolic abstractions to guide the inference process towards a limited range of possible and valid conclusions.

Qualitative abstractions of probabilities include: ranking based on infinitesimal probabilities (or $\in$-semantics) where order of magnitudes of a small quantity ($\in$) are used to rank outcomes based on their likelihood (Goldszmidt & Pearl, 1996). Another approach is to assess the relative changes in probabilities as either +, − , 0, or ? to indicate an increase or a decrease, no change or unknown in a posterior probability given some evidence (Parsons, 1995). Other approaches are rooted in a probabilistic abstraction of defaults as in (Neufeld, 1989). For a more complete review refer to (Parsons, 2001) and (Wellman, 1994).

These approaches to qualitative probabilities do not lend themselves to temporal reasoning. The infinitesimal probabilities assume shallow reasoning chains and would not capture the accumulation of small changes over time (Goldszmidt & Pearl, 1996). For example, if the probability of death increases by a constant $\in$ each time unit that a given infection is left untreated, the order of magnitude does not change and an imminent certain danger would be missed.

Qualitative techniques based on relative change in probabilities rely on expressing the direction of change in posterior probabilities of a hypothesis given some evidence. The type of question that such representation answers is how an observation (e.g. fever) would change belief in a hypothesis (e.g. flu)? Would it increase that belief or decrease it? These changes are therefore measured with respect to an assumed prior belief. In temporal reasoning, it is important to express the change in beliefs with respect to previous (or future) time units. These beliefs change even in the absence of evidence. For example, it is important to be able to represent the fact that as time passes colors fade. It is also necessary to find a concise representation for efficient acquisition. It is not acceptable to represent the probability that color fades given each possible period. Relative change techniques also suffer from the spreading of unknown values when used in long inference chains.

The following section introduces the qualitative probability model. Inference using these qualitative probabilities relies on enforcing probabilistic and temporal constraints implied by the laws of probabilities and temporal precedence ordering. Sections 3 and 4 introduce these constraints. Section 5 discusses how to apply the qualitative probability model to inference tasks.

2. The Qualitative Probabilities

To express change in probability over time, belief increases (and decreases) ought to be measured with respect to time. These changes are referred to, in this work, as time-relative changes. Similarly, it is possible to introduce time-infinitesimal probabilities because belief increases (and decreases) correspond to infinitesimal changes over infinitesimally short periods.

A combination of time-relative change and time-infinitesimal probabilities is used. The value of a qualitative probability can be increasing *(i)*, decreasing *(d)*, constant *(c)* or unknown *(?)*. These values are defined in terms of inequalities first, then presented in terms of infinitesimal probabilities as follows:

$P_{[t\text{-}t+\Delta]}(x)=d \Rightarrow \forall\eta,\tau{:}\eta<\tau<\Delta,\ P_{t+\eta}(x)>P_{t+\tau}(x),$

$P_{[t\text{-}t+\Delta]}(x)=i \Rightarrow \forall\eta,\tau{:}\eta<\tau<\Delta,\ P_{t+\eta}(x)<P_{t+\tau}(x),$

$P_{[t\text{-}t+\Delta]}(x)=c \Rightarrow \forall\eta,\tau{:}\eta<\tau<\Delta,\ P_{t+\eta}(x)=P_{t+\tau}(x),$

$P_{[t\text{-}t+\Delta]}(x)=? \Rightarrow \forall\eta,\tau{:}\eta<\tau<\Delta,\ P_{t+\eta}(x)\gtreqless P_{t+\tau}(x).$

In the above, η and τ are time points within the interval [t, t+Δ].

The ϵ -semantic for d, i, c, and ? are defined as follows:

$P_{[t\text{-}t+\Delta]}(x)= d \Rightarrow \forall\eta,\tau{:}\eta<\tau<\Delta,\ \exists\,\epsilon$ such that

$P_{t+\tau}(x)-P_{t+\eta}(x)=-\epsilon,$

$P_{[t\text{-}t+\Delta]}(x)=i \Rightarrow \forall\eta,\tau{:}\eta<\tau<\Delta,\ \exists\epsilon$ such that $P_{t+\tau}(x)-P_{t+\eta}(x)=\epsilon,$

$P_{[t\text{-}t+\Delta]}(x)=c \Rightarrow \forall\eta,\tau{:}\eta<\tau<\Delta\ \exists\epsilon$ such that $P_{t+\tau}(x)-P_{t+\eta}(x)|=0,$

and $P_{[t\text{-}t+\Delta]}(x)=? \Rightarrow \forall\eta,\tau{:}\eta<\tau<\Delta,\ \exists\epsilon$ such that

$P_{t+\tau}(x)-P_{t+\eta}(x)$ is in the interval $\{[-\epsilon,\epsilon]\}$.

In the above definitions, both delta and epsilon are small positive quantities.

Any probability *P(x)* (or probability density) can be expressed as a function in ϵ and *t*. The sign of the of time-slope (differential with respect to time) of this function at a given *t* corresponds to the time-change probability at that time.

To represent how a probability or a probability density function (pdf) changes over time, a string of literals is used such that each literal corresponds to a qualitative probability.

These strings cannot have two adjacent identical literals. Therefore a Gaussian distribution is "*id*", a uniform distribution is "*c*", and so on. Subscripts can be added to a literal to indicate the time interval for the literal. For example, a Gaussian distribution of mean w can be described as "$i_{[-inf,w]}d_{]w,inf]}$". It is preferable however to use finite intervals in describing distributions for practical reasons, therefore the string "$i_{[0,w]}d_{]w,2w]}$" approximates the Gaussian distribution. The notation described in this section is used throughout the paper to describe different functions of time. The values ϵ, and $1\text{-}\epsilon$ denote a low probability value and a high probability value respectively.

3. Probabilistic Constraints

The constraints implied by the axioms of probabilities include constraints on probability density functions and constraints on probability values.

3.1 Constraints on Probability Densities

Lemma 1 Any probability density function ending with "*i*" or "*c*" can only be defined on a finite interval. Formally, if "*x*i*" or "*x*c*" is a probability density function and *x** represent zero or more literal in *{d,c,?}*, then there exist a finite $w<\infty$ such that the density function is not defined beyond *w*.

Proof. An increasing or constant density function[1] should be bound so that the probability does not exceed unity.

[1] This result is not true for hazard functions representing the rate of occurrence of an event.

Lemma 2 Chopping off the tail of any probability density function ending with "*d*" introduces an arbitrarily small error ϵ if the tail is chopped beyond *w*. Formally, if "*x*d*" is a probability density function and *x** represent zero or more literal in *{i,c,?}*, then there exist a finite $w<\infty$ such that any probability is affected by a finite amount ϵ if the distribution beyond *w* is ignored.

Proof. The contribution of a decreasing density function possibly extending till infinity decreases as it approaches infinity such that the contribution of the tail can be bound to an arbitrarily small error by choosing an appropriate cutoff point *w*.

Proposition 1 There is a finite time approximation for any probability density function that introduces a small error of at most ϵ in a probability evaluation.

Proof. This proposition follows from Lemma 1 and Lemma 2 above assuming that all temporal probability densities are defined over the positive time line and that any such function must end with "*c*", "*i*" or "*d*".

3.2 Probabilistic Inference Rules

Proposition 2 The probability of the complement of *x* over a given time interval, *P(~x)* can be evaluated based on the probability of *x*. Let *P(x)* denote the change in the probability of *x* over some interval and *P(~x)* denote the change in *~x* probability for the same interval. *P(x)* and *P(~x)* are related by the following relation:

P(x)	***P(~x)***
I	*d*
d	*I*
c	*c*
?	*?*

Proof. The above table follows directly from the notion that *P(~x) = 1- P(x)*.

Proposition 3 The following table gives the product of two probabilities:

*	***I***	***d***	***c***	***?***
I	*I*	*?*	*i*	*?*
d	*?*	*d*	*d*	*?*
c	*I*	*d*	*c*	*?*
?	*?*	*?*	*?*	*?*

Proof. The above table preserves the slope of a probability when multiplied by a constant or another probabilities of the same slope. The product of two probabilities of different slopes is unknown and so is the product involving an unknown slope.

Bayes Rule: Applying the above product proposition, we can calculate a joint probability. According to Bayes rule, the joint probability is the product of a conditional probability and a prior probability.

Assume that the joint probability *P(A,B)= i*. According to Bayes Rule, *P(A,B) = P(A|B) P(B) = P(B|A) P(A).* Therefore, having *P(A,B)=i* is inconsistent in the above table with having *P(B) =d and P(A|B) = d*.

Proposition 4 The table below represents the sum of two probabilities.

+	***i***	***d***	***c***	***?***
i	*i*	*?*	*i*	*?*
d	*?*	*d*	*d*	*?*
c	*i*	*d*	*c*	*?*
?	*?*	*?*	*?*	*?*

3.3 $\in$ -Semantics Identities

Lemma 3 Both $\in$ and $(1-\in)$ converge to zero when raised to a sufficiently large exponent. Mathematically,

$$Lim_{n\to\infty}\ \in^n=0 \text{ , and } Lim_{n\to\infty}\ (1-\in)^n=0.$$

Lemma 4 The sum of powers of $\in$ converges to a constant that is less than or equal to 1.

$$\Sigma_i \in^i = c \preceq 1.$$

Proof. This follows directly from Lemma 3 and from the fact that these values form a probability distribution, as the added terms approach zero, the total approaches a constant. This constant must be bound by 1 to be a probability.

Proposition 5 To achieve the desirable accumulation of small effects, it is assumed that the sum of a sufficiently large number of small probabilities may add up to 1.

$$Lim_{n\to\infty}\ n\in \preceq 1.$$

Proposition 6 It follows from proposition 5 that the sum of a large number of $\in^m$ approaches $\in^{m-1}$.

$$Lim_{n\to\infty}\ n\in^m \preceq \in^{m-1}$$

4. Temporal Constraints

It is useful to distinguish between hard temporal constraints and soft temporal constraints. As usual hard constraints must hold while soft constraints may be violated. In a sense, soft constraints are similar to defaults that could be assumed in the absence of contradictory evidence. For example, statements like “I eat a light meal before going to work” or “Students take data structures before artificial intelligence” represent soft constraints. Hard constraints on the other hand represent physical laws that are not to be violated. As an example of hard constraints, consider the statement “An object cannot be in more than one location at the same time.” According to multiple worlds semantics, a state is that is not reachable from another state represent a hard constraint, while a soft constraint represents a world state that is the not most normal or preferred.

The distinction between hard and soft constraints has implications on the representation. Soft constraints have a probability associated with them indicating how likely they are to be satisfied. Any scenario that violated hard constraints is rejected. However, scenarios violating soft constraints may be unlikely. Both hard and soft constraints may be temporal or static. Temporal constraints impose a precedence relation between two or more time points while static constraints apply to a set of states at a particular time. It is necessary to introduce the time model before showing how to represent temporal constraints.

4.1 Time Model

A time state is described as a pair < t, S > where t is a time point, duration or interval measured from an arbitrary origin and S is a state. Any state S generally consists of a conjunction of assignments of values to state variables. Each time t has a set of possible states associated with it. These states may share some values but any two states will differ in at least one value assignment. Each possible state occurs with a certain probability.

There are sound reasons to support the view that the past is unique. However, this unique past is rarely known. Instead of assuming a unique past, allowing for uncertainty about states in the past would result in a set of possible pasts as well as a set of possible presents and futures.

A chronicle is a self-consistent history with a past, present and a future. Different chronicles may share states and must share observed variable assignment (assuming reliable observations). A chronicle is self-consistent if all states do not violate hard constraints, and that all state change implied are possible (i.e. may result from known and possible events, actions, persistence or natural change).

4.2 Representing Temporal Constraints

Hard temporal ordering constraints: I and J are two time-state structures, H is a hard ordering constraint in the form HO: M(R(I, J)) where M is a temporal modal operator, and R is a temporal relationship between I and J. Therefore, any chronicle that does not satisfy R is not acceptable at some or all times as specified by M. The use of the modal operators (always, until, eventually and next) in expressing constraints is useful in making distinction between hard constraints that must eventually hold from those that must hold always or within a certain time horizon. The relationship R can be any of Allen’ s interval relationships (Allen , 1984).

Hard constraints on temporal probabilities: These constraints are expressed as HT:M(P(I)=p), where M is a temporal modal operator, P(I) is the probability of time-state I, and p is in {d, i , c}. This allows us to express something like “eventually the probability of a valve failure increases”.

HT:◇ (P(valve_failure) =i) where ◇ denotes an eventuality. Therefore, only chronicles where this probability eventually increases are acceptable.

Note that it is possible to have combined temporal ordering and temporal probability constraints or HOT constraints. For example, stating that the probability of secondary infection increases following a transplant would require both a temporal ordering constraint (after) and a temporal probability constraint (increasing probability of infection).

Soft temporal ordering constraints: Normal orderings are expressed as soft ordering constraints. A chronicle that satisfies these constraints is preferred. However, there is a probability that these constraints will be violated. For example, students normally take a data structures course before taking an artificial intelligence course would be represented by such a constraint. This statement is represented as SO: before(data_structures, ai).

Soft constraints on temporal probabilities: This class of constraints captures normal changes in probabilities. To state that the probability of precipitation usually increases as we approach the rainy season would be represented by:
ST: P_J(precipitation) =i where J is the time interval from now to the rainy season. Again chronicles consistent with this statement are preferred to others.

In general, it is possible to combine soft ordering and probability constraints. It is also possible to have some hard constraints apply when soft constraints are holding (or not holding). For example, if students who did not take the data structures course before artificial intelligence are required to take it as a co-requisite, we end up with a hard constraint having to hold whenever the soft constraint is violated.

5. Inference Mechanisms

The qualitative probability framework presented so far can be applied in a variety of ways to carry out different reasoning tasks. The qualitative probability model imposes constraints that allow the reasoning engine to rule out a set of inconsistent scenarios. The following subsections illustrate how apply these constraints. The first subsection shows how to rule out inconsistent probability distributions. The second subsection gives a qualitative formulation for persistence and convergence to a stationary distribution. The third subsection overviews a technique for causal reasoning without expanding all possible chronicle trees.

5.1 Probabilistic Consistency

Survival analysis techniques have been used in probabilistic temporal reasoning (Tawfik & Neufeld, 2000). In survival analysis, a failure density function represents the pdf of failure as a function of time. Another somewhat related function, the hazard function, represents the instantaneous rate of occurrence of failure. In other words, the hazard function at time T represents the portion that failed at time T after surviving till then. The following table from (Sibuya, 1996) shows that out of 36 hazard – failure combinations, the 16 marked by * are impossible.

h/f	*I*	*d*	*id*	*di*	*idi*	*did*
I						
d	*		*	*	*	*
id	*			*	*	
di	*		*		*	
idi	*					
did	*		*	*	*	

Proof

We have a hard constraint stating that $h(t)= f(t) / S(t)$ where $S(t) = 1 - \int f(t)\,dt = d$.

For $f(t)=i$, and $S(t)=d$. The ratio $i/d = i$ necessarily. Therefore, the only possible $h(t)$ for $f(t) = i$ is also i and the other 5 options in the first column are impossible.

Given that $f(t) = h(t)\ S(t)$, and that $S(t)=d$ always, then if $f(t) = d$ then $h(t) = c$ or d, according to Proposition 3. Therefore, all the options in row 2 of the above table are impossible except d. For $h(t) = id$, $f(t) = id*d = ?d$. Therefore, it is impossible to have f(t) =di or idi. For $h(t)=di$, $f(t) = di*d= d?$, therefore it is impossible to have $f(t) = id$ or *idi*. In the last row, we have $h(t) = did$, then $f(t)=did*d = d?d$. Therefore, it is impossible to have $f(t)=id$, *di* or *idi*.

Another useful result follows from Lemma 1, all the density functions ending with *i* (*i*, *di*, and *idi*) are defined over a finite time range.

5.2 Persistence and convergence

This subsection shows that the qualitative probabilities presented capture many fundamental notions in probabilistic temporal reasoning. The first of these notions is persistence (i.e. the tendency of a fluent to remain unchanged). The second is that of indigenous and exogenous change. The third notion is convergence to a particular state (absorption) or convergence to a particular probability distribution.

A binary fluent is subject to its persistence, indigenous and exogenous factors forcing it to become false, and similar factors forcing it to become true.

Persistence is the probability that a fluent maintains its truth for a period of time. Persistence depends on the probabilities of change (Tawfik & Neufeld, 2000). For a fluent (f) to persist for any short duration Δ, the probability $P(f_{t+\Delta}|f_t)$ must be greater than $\in$. By doubling the duration, $P(f_{t+2\Delta}|f_t)= P(f_{t+2\Delta}|f_{t+\Delta})\ P(f_{t+\Delta}|f_t)+ P(f_{t+2\Delta}|\sim f_{t+\Delta})P(\sim f_{t+\Delta}|f_t)$. However, the second term does not represent persistence but rather two opposite changes. Therefore the persistence is given by $\Pi_i P(f_{t+i\Delta}|f_{t+(i-1)\Delta})$. As the duration grows longer, the probability of persistence approaches 0 according to Lemma 3 (if $P(f_{t+\Delta}|f_t)\neq 1$). In general, persistence is limited.

The remaining probability mass ($1- P(f_{t+\Delta}|f_t)$) represents the probability of change during a duration Δ. A binary fluent can change from true to false and from false to true. First, consider a fluent that can only change from true to false. The probability $P(\sim f_{t+\Delta}\ |f_t) = 1- P(f_{t+\Delta}\ |f_t)$, and after 2 Δ, $P(f_{t+2\Delta}|f_t) =1- P(f_{t+2\Delta}|f_{t+\Delta})\ P(f_{t+\Delta}|f_t)$, and as the probability of persistence approaches 0, this probability approaches 1. This represents a case of absorption because, f cannot become true again.

In the general case, both transitions from true to false and from false to true are allowed. In such cases, by solving a recurrence relation as presented in (Tawfik & Neufeld, 2000), it is possible to obtain an expression for the probability $P(f_{t+n\Delta}|f_t)$ that depends on the transition probabilities $P(f_{t+\Delta}\ |\sim f_t)=a$ and $P(\sim f_{t+\Delta}\ |f_t)=b$.

$$P(f_{t+n\Delta}\ |f_t) = a(1- a - b)^{n-1}[(1/(a + b)) - 1]+ b / (a + b).$$

The term $(1- a - b)^{n-1}$ represents the probability of persistence. This probability decreases over time and for a sufficiently large n, as n approaches infinity, and the above expression converges to a constant $b/(a + b)$ that reflect the relative forces of change.

5.3 Deducing the Effects of Events

A rather inefficient way for temporal uncertain reasoning is based on chronicle trees. The idea is to expand chronicle trees such that all evidence states are common to all chronicles. Each chronicle is checked for consistency with hard constraints. Chronicles that are not consistent with these constraints are rejected. Chronicles are ranked based on how well they satisfy the soft constraints. The expansion of the chronicle trees should be exhaustive to allow all possible times, events, and actions. This is the main source of inefficiency.

However, a careful examination of the chronicle trees reveals that to a large extent temporal projections carried out using numeric convolution provide a concise way for evaluating ranges of event or action intervals and their consequences. In fact a convolution examines all possible effect times following each possible event time. The use of convolution in temporal reasoning has been introduced in (Dean & Kanazawa, 1989) and re-examined in (Tawfik & Neufeld, 2000). In using convolution, one has to assume that the probability distribution of the time of occurrence of the event (or action) is given as well as an distribution representing the time required for the effect to materialize, and that these two distributions are independent. The convolution gives the distribution of the effect. Using the time-change qualitative probabilities presented here in performing convolutions gave some interesting and promising results. The first result is that the distribution of the effect of an event took the shape icd or id for many event and effect distributions. The reason for these shapes is that for two finite continuous distributions, the value of the integral increases gradually as the two distributions meet, it reaches a maximum, then decreases. The second result is that duration of the resulting distribution is equal to the sum of the two distributions. These two results simplify the exercise of evaluating effects considerably. Temporal ordering consistency tests is done using temporal constraint satisfaction.

6. Conclusions

The qualitative probabilities presented here provide a promising alternative particularly attractive for temporal reasoning. The relative change abstraction differs in other change-based qualitative probabilities in that the changes are expressed over time rather than relative to a prior. The infinitesimal approximation presented does not use order of magnitude ranking and allows the accumulation of effects over time. This work has also presented a number of useful and important results that proved to be rather easy within this framework. Some of the main results are:

- A finite time probability can always be used to approximate an infinite distribution.
- It is possible to use qualitative probabilities to impose probabilistic and temporal constraints.
- Efficient evaluation of effect distribution can be done using geometric convolution.

Acknowledgements

The author would like to thank the reviewers for their helpful comments and to acknowledge the support of the Natural Sciences and Engineering Research Council (NSERC) and the University of Windsor.

References

Allen, J. F. 1984. Towards a General Theory of Action and Time. Artificial Intelligence. 23, 123-154.

Dean, T. and Kanazawa, K. 1989. A Model for Reasoning about Persistence and Causation. Computational Intelligence. 5(3) 142-150.

Goldszmidt, M., & Pearl, J. 1996. Qualitative Probabilities for Default Reasoning, Belief Revision and Causal Modeling. Artificial Intelligence, 84, 57-112.

Neufeld, E. 1989. Defaults and Probabilities: Extensions and Coherence. First International Conference on the Principles of Knowledge Representation and Reasoning, Toronto, Morgan Kaufmann.

Parsons, S. 1995. Further results in qualitative uncertainty. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 3(1), 1-24.

Parsons, S. 2001, Qualitative Methods for Reasoning under Uncertainty, MIT Press, Cambridge, MA.

Sibuya, M. 1996. The shapes of a probability density function and its hazard function. In Jewell et al. Lifetime Data: Models in Reliability and Survival Analysis (307-314). Netherlands: Kluwer Academic Publishers.

Tawfik, A.Y., & Neufeld, E. 2000. Temporal Reasoning and Bayesian Networks. Computational Intelligence, 16(3), 349-377.

Wellman, M.P. 1994. Some Varieties of Qualitative Probability. Fifth International Conference on Information Processing and Management of Uncertainty in Knowledge-based systems.