APPROXIMATE INFERENCE AND FORECAST ALGORITHMS IN GRAPHICAL MODELS FOR PARTIALLY OBSERVED DYNAMIC SYSTEMS

Alberto Lekuona, Beatriz Lacruz, Pilar Lasala
Departamento de Métodos Estadísticos
Universidad de Zaragoza (Spain)
e-mail: {lekuona,lacruz,lasala}@posta.unizar.es

ABSTRACT

From a statistical point of view, modelling stochastic temporal processes by graphical models is a suitable choice, specially when certain standard assumptions in classical modelling cannot be assumed.

Focusing the discussion on partially observed domains, it is important to design algorithms which provide probability distributions over the current and future states of the non-observable components of the domain, using the information stored in the observable components.

In this paper, we present a simulation algorithm for approximating the exact probability distributions associated with such inference and forecast processes. This algorithm uses both the probabilities built at the previous time step and the new evidence obtained to propose new probability distributions associated with current and future states of the domain. To validate the algorithm, a case study of equipment maintenance is considered.

1 INTRODUCTION

Generally speaking, a dynamic graphical model may be defined as a sequence of graphical submodels, each representing the state of a system at a particular time slice, interconnected by temporal relationships. These temporal relationships stand for the dependencies between the state of the system in a certain time slice and its state in later time slices and capture the dynamic behaviour of the domain variables. Finally, the dynamic graphical model as a whole represents the evolution of a dynamic system over the time.

Dynamic graphical models is an outstanding research topic and several authors have proposed several solutions [Berzuini,90], [Kjærulff,92], [Dagum,Galper,94], [Provan,94], [Kanazawa et al.,95], [Hanks et al.,95]. Most of the previous approximations are constrained to directed acyclic graphs and some of the imposed conditions are, in general, too restrictive. For these reasons, it seems that an alternative graphical model and algorithms for inference and forecasting associated with it should be considered.

In [Lekuona et al.,96] a new dynamic graphical model for modelling partially observed dynamic systems and exact algorithms for inference and forecasting are proposed. Following with this model, in this paper we introduce an approximate algorithm for inference and forecasting using stochastic simulation. This algorithm is validated with a case study.

This paper is organized as follows: in sections 2 and 3, we describe the graphical model to be considered and the algorithm for making exact inferences and forecasts. Based on this algorithm, in section 4 we present an approximate algorithm for these two tasks. Finally, in section 5 we consider an optimal stopping example to validate the proposed algorithm.

2 MODELLING PARTIALLY OBSERVED DYNAMIC SYSTEMS

Let $\{X_t\}_{t\in\Gamma}$ be a stochastic process ($\Gamma \subseteq \mathbb{N}$ and $X_t = (X_t^1, \ldots, X_t^n)$), describing the changes in the state of a certain system over time.

The definition of a *graphical structure associated with* $\{X_t\}_{t\in\Gamma}$, denoted by $G_{\{X_t\}_{t\in\Gamma}}$, is based on the union of a sequence of chain graphs $\{G_t\}_{t\in\Gamma}$, which describes relationships among the components of X_t in any time slice t, and a sequence of directed acyclic graphs $\{G_{kt}\}_{\substack{k,t\in\Gamma \\ k<t}}$ which describes non contemporary relationships among a time slice t and previous time slices $k < t$. The former stands for the system state description in any time slice and the latter stands for the dynamic behaviour of the system. Figure 1 shows an example of a graphical structure associated with a dynamic system, depicted in three time slices.

Given a dynamic system described by a stochastic process $\{X_t\}_{t\in\Gamma}$, we define *a dynamic graphical model* as a tern $\left(\{X_t\}_{t\in\Gamma}, G_{\{X_t\}_{t\in\Gamma}}, \{P_t\}_{t\in\Gamma}\right)$ where $G_{\{X_t\}_{t\in\Gamma}}$ stands for a graphical structure associated with $\{X_t\}_{t\in\Gamma}$ and $\{P_t\}_{t\in\Gamma}$ is a sequence of probability distributions where P_t is $\bigcup_{k\leq t}\left[G_k \bigcup \left(\bigcup_{k'<k} G_{k'k}\right)\right]$-markovian, $\forall t \in \Gamma$.

If t is fixed, the graph $G_{0t} = \bigcup_{k\leq t}\left[G_k \bigcup \left(\bigcup_{k'<k} G_{k'k}\right)\right]$ turns out to be a chain graph and represents the whole qualitative structure of the model from the first time slice to the current time t, and therefore, it summarizes all possible influences of the past in the current time slice. This graph is associated with a probability measure P_t which is coherent with it, in the markovian sense.

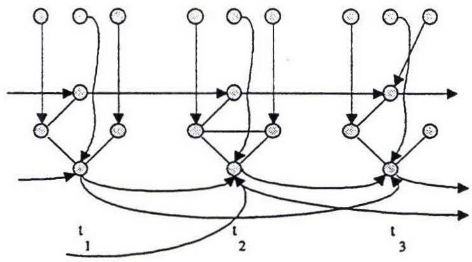

Figure 1

Furthermore, note that the previous definition generalizes the concept of graphical model associated with a static domain in a natural way.

If the system to be modelled is partially observed, variables $(X_t^1, \ldots, X_t^n)$ can be classified as *observable variables* $E_t = \{E_{it}, i = 1, \ldots, k\}$, and *non observable variables* $H_t = \{H_{jt}, j = 1, \ldots, m\}$, $k + m = n$. In what follows, we assume that observable variables will always be observed and their evidence will be instantaneously obtained.

The main goal in modelling these partially observed domains is to make inferences about the states of the non observable variables in a certain time slice, given a specific evidence accumulated up to that time slice. In the same way, building a forecast process of the future states can be appealing.

In the graphical structure associated with the process $\{X_t\}_{t\in\Gamma}$, two types of relationships among variables will be considered:

contemporary relationships, which will be undirected between non observable variables (there is no reason to describe relationships between non observable variables, with more restrictive edges; furthermore, in this way, there is a standard handling of the components of H_t) and directed from observable to non observable variables (these edges mean that evidence collected in the current time modifies our beliefs over the states of the non observable variables); and *non contemporary relationships*, which will be directed and only between non observable variables (these relationships represent the evolution of the non observable components over time and they will cooperate with directed contemporary relationships).

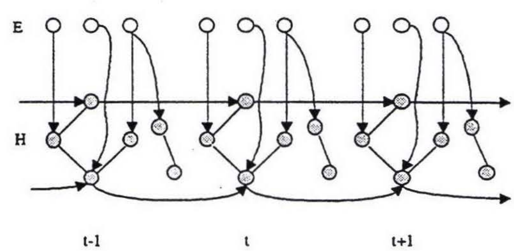

Figure 2

All this is shown in Figure 2. The possible remaining relationships will not be considered since we want a parsimonious model and are interested in making inferences and forecasts for the non observable variables. The considered relationships are sufficient for carrying out these tasks.

To model the whole system, we need a stochastic model for the non observable variables. We focus the discussion on a markovian behaviour of the process $\{H_t\}_{t\in\Gamma}$:

$$H_{t_n} \perp H_{t_0}, \ldots, H_{t_{n-2}} | H_{t_{n-1}} \quad [P] \qquad \forall t_0 < t_1 < \cdots < t_n$$

where P is a probability distribution properly defined in some product space. With this condition, non contemporary relationships between every two consecutive time slices have to be fixed and there will not simultaneously be relationships from two time slices to later time slices.

Considering the chain graph G_{0T}, $T > 0$ and given the strictly positive probability distributions $P_t(E_{it})$ and $P_t(\tau_{jt}|bd(\tau_{jt}))$, $i = 1, \ldots, k$, $j = 1, \ldots, h$ and $t = 0, \ldots, T$, where τ_{jt} stands for the chain components in G_{0T} and $(P_t)_{cl(\tau_{jt})}$ is $\left(G_{0T\,cl(\tau_{jt})}\right)^m$-markovian $\forall j, t$, it can be shown that $\left(\{E_{1t}, \ldots, E_{kt}, H_{1t}, \ldots, H_{mt}\}_{t\in\Gamma}, \{G_{0t}\}_{t\in\Gamma}, \{P_{0t}\}_{t\in\Gamma}\right)$ is a dynamic graphical model where

$$P_{0T}\left(E_{10},\ldots,E_{k0},H_{10},\ldots,H_{m0},\ldots,E_{1T},\ldots,E_{kT},H_{1T},\ldots,H_{mT}\right)=$$
$$=\prod_{t=0}^{T}\prod_{i=1}^{k}P_t\left(E_{it}\right)\prod_{j=1}^{h}P_t\left(\tau_{jt}|bd\left(\tau_{jt}\right)\right)$$

Therefore, marginal probabilities of each observable component E_{it} and probabilities of each chain component in G_{0T}, conditional to its boundary, are needed to built P_{0T}. These latter probabilities collect both non observable components in the previous time slice and the influence of contemporary observable components.

As a consequence of the above result and considering the graph in Figure 3, denoted by $G_{\{(E_t,H_t)\}_{t\in\Gamma}}$, we obtain that

$$\left(\{(E_t,H_t)\}_{t\in\Gamma}, G_{\{(E_t,H_t)\}_{t\in\Gamma}}, \{P_{0t}\}_{t\in\Gamma}\right)$$

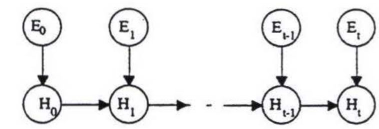

Figure 3

is also a dynamic graphical model, with

$$P_{0T}\left(E_0,H_0,\ldots,E_T,H_T\right)=\prod_{t=0}^{T}P_t\left(E_t\right)P_t\left(H_t|H_{t-1},E_t\right)$$

3 EXACT INFERENCE AND FORECAST ALGORITHMS

The main goal is to obtain probability distributions to measure the current and future states of the non observable variables given the information stored in the observable variables, i.e., to obtain $P_{0t}\left(H_t|D_t\right)$ and $P_{0t}\left(H_{t+1}|D_t\right)$ $(t \geq 0)$, where D_t stands for all the information known at time t.

To calculate $P_{0t}\left(\bullet|D_t\right)$, the probability distributions $P_s\left(\tau_{js}|bd\left(\tau_{js}\right)\right)$ and $P_s\left(E_{is}\right)$ $(j=1,\ldots,h,\ i=1,\ldots,k,\ s\leq t)$, are supposed to be known.

These distributions will be given by an expert or will be automatically constructed out of some suitable sampling information.

Exact algorithms for inference and forecasting are discussed in [Lekuona et al.,96]. According to Figures 4 (inference) and 5 (one-step forecast), where shaded nodes stand for observed variables, the formulas to be used are:

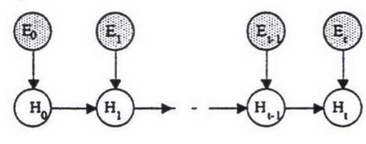

Figure 4

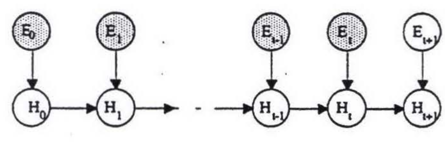

Figure 5

Step 1 (Inference in t=0). *Measure observable variables $E_0 = e_0$ and consider D_0; compute*

$$P_{00}\left(H_0|D_0\right)=\prod_{j=1}^{h}P_{00}\left(\tau_{j0}|bd\left(\tau_{j0}\right)\right)$$

Step 2 (Forecast in t=1). *Compute*

$$P_{00}\left(H_1|D_0\right)=\sum_{E_1}\prod_{i=1}^{k}P_{01}\left(E_{i1}\right)\left(\sum_{H_0}\prod_{j=1}^{h}P_{01}\left(\tau_{j1}|bd\left(\tau_{j1}\right)\right)P_{00}\left(H_0|D_0\right)\right)$$

Step 3. *Stop or let $t = t + 1$ and go to Step 4.*

Step 4 (Inference). *Measure observable variables $E_t = e_t$ and consider D_t; compute*

$$P_{0t}\left(H_t|D_t\right)=\sum_{H_{t-1}}\prod_{j=1}^{h}P_{0t}\left(\tau_{jt}|bd\left(\tau_{jt}\right)\right)P_{0t-1}\left(H_{t-1}|D_{t-1}\right)$$

Step 5 (One-step forecast). *Compute*

$$P_{0t}\left(H_{t+1}|D_t\right) = \sum_{E_{t+1}} \prod_{i=1}^{k} P_{0t+1}\left(E_{it+1}\right) \left(\sum_{H_t} \prod_{j=1}^{h} P_{0t+1}\left(\tau_{jt+1}|bd\left(\tau_{jt+1}\right)\right) P_{0t}\left(H_t|D_t\right) \right)$$

Step 6. *Go to Step 3.*

Note that these expressions are built iteratively. In each inference step, beliefs over the states at time t are generated from beliefs generated at time $t-1$ and the new information received. In each forecasting step, beliefs over the states at time $t+1$ are generated from a weighted mean of the inference obtained when E_{t+1} is measured. The weights are just the probabilities of these measures.

4 AN APPROXIMATE INFERENCE AND FORECAST ALGORITHM

It is clear that stochastic simulation algorithms, which provide approximations to the exact probabilities, are needed as the number of variables grows.

For purpose of inference and forecasting and making use of the features of the graphical model we propose, these tasks are well represented by the graphs in Figure 6.

Based on the formulas in Section 3, the proposed algorithm for approximate inference and forecasting is the following:

Step 1 (Inference in t=0). *Measure observable variables $E_0 = e_0$ and consider D_0; compute*

$$P_{00}\left(H_0|D_0\right) = \prod_{j=1}^{h} P_{00}\left(\tau_{j0}|bd\left(\tau_{j0}\right)\right)$$

Step 2 (Forecast in t=1). *Estimate $P_{00}\left(H_1|D_0\right)$ using simulation techniques.*

Step 3. *Stop or let $t = t+1$ and go to Step 4.*

Step 4 (Inference). *Measure observable variables $E_t = e_t$ and consider D_t; estimate $P_{0t}\left(H_t|D_t\right)$ according to Figure 6a, using simulation techniques.*

Step 5 (One-step forecast). *Forecast states of non observable variables in the time slice $t+1$ given D_t, estimating $P_{0t}\left(H_{t+1}|D_t\right)$ by simulation techniques according to Figure 6b.*

Step 6. *Go to Step 3.*

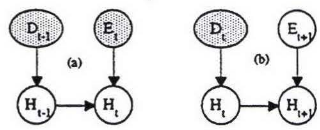

Figure 6

The procedures to simulate $P_{0t}\left(H_t|D_t\right)$ and $P_{0t}\left(H_{t+1}|D_t\right)$ can be:

PROCEDURE to simulate $P_{0t}\left(H_t|D_t\right)$
loop for $i = 1, \ldots, N$
Instantiate $H_{t-1} = h_{t-1}$ using $P_{0t-1}\left(H_{t-1}|D_{t-1}\right)$
Instantiate $H_t = h_t$ using $P_{0t}\left(H_t|H_{t-1} = h_{t-1}, E_t = e_t\right)$
Add sample $H_t = h_t$ in a frequency table
end loop
$P_{0t}\left(H_t = h_t|D_t\right) := fr\left(H_t = h_t|D_t\right), \forall h_t$

PROCEDURE to simulate $P_{0t}\left(H_{t+1}|D_t\right)$
loop for $i = 1, \ldots, N$
Instantiate $H_t = h_t$ using $P_{0t}\left(H_t|D_t\right)$
Instantiate $E_{t+1} = e_{t+1}$ using $P_{0t+1}\left(E_{t+1}\right)$
Instantiate $H_{t+1} = h_{t+1}$ using $P_{0t+1}\left(H_{t+1}|H_t = h_t, E_{t+1} = e_{t+1}\right)$
Add sample $H_{t+1} = h_{t+1}$ in a frequency table
end loop
$P_{0t}\left(H_{t+1} = h_{t+1}|D_t\right) := fr\left(H_{t+1} = h_{t+1}|D_t\right), \forall h_{t+1}$

5 A CASE STUDY

The case study is based on the *equipment maintenance and replacement problem* which is described, in general terms, as follows:

A decision maker periodically inspects the conditions of an equipment which produces an item. The equipment deteriorates with time and the decisor, after the inspection, decides on the extent of maintenance, if any, to carry out. Choices may vary from routine maintenance to equipment replacement.

The evolution of the *unknown* real state of the equipment is modelled by a Markov chain $\{X_t\}$ with both initial probability distribution and non homogeneous transition probability matrix unknown. In order to simplify the model the assumption of ten possible working states and the sequential deterioration of the system is imposed. Sequential deterioration means that, if the system is working in state j at time t, it will work in state j or $j-1$ at time $t+1$. State 1 stands for the absorbing state corresponding with the worst working state. Therefore, the transition probability matrix from time t to time $t+1$ is as follows:

$$P_{t,t+1} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1-\theta_2^t & \theta_2^t & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1-\theta_M^t & \theta_M^t \end{bmatrix}$$

where θ_i^t stands for the probability of remaining in the state i in the state evolution from t to $t+1$. These θ_i^t are the non observable variables in the graphical model, depicted in Figure 7, together with the initial state of the system, described by the node X_0. The purpose in this graphical modelling is to estimate these θ_i^t in any time slice t.

The observable variables $\{Y_t\}$, which measure the percentage of non defective items made by time period and are categorized in ten possible classes, summarize the equipment inspection. The dynamic graphical structure associated with this problem is shown in Figure 7.

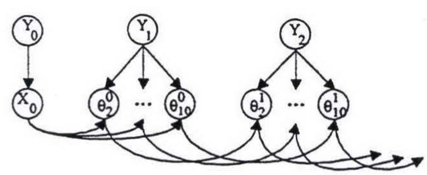

Figure 7

The sequential process to be described by the graphical model consists of the next steps:

1 Consider D_0. Infer X_0 through $P(X_0|D_0)$;

2 Do $t = t+1$. Consider D_t. Build the conditional distributions $\theta_2^{t-1}|D_t, \ldots \theta_M^{t-1}|D_t$;

3 Using Bayes estimation, estimate $\theta_2^{t-1}, \ldots, \theta_M^{t-1}$. Infer X_t through the transition equation

$$P(X_t|D_t) = P(X_{t-1}|D_{t-1})\,\hat{P}_{t-1,t}$$

where $\hat{P}_{t-1,t}$ stands for the estimated transition probability matrix from time $t-1$ to time t, associated with $\{X_t\}$;

4 Decide the action to take. If continue, go to 2. Else, stop.

For this process, the probabilities $P_t(Y_t)$, $P_0(X_0|Y_0)$, $P_1(\theta_i^0|X_0, Y_1)$ and $P_t(\theta_i^{t-1}|\theta_i^{t-2}, Y_t)$ have to be known. In order to check the behaviour of the algorithm, these probabilities have been modelled using Normal and Beta distributions (as [Silver,63] proposes) and linear transformations of them. The proposed Bayes estimator for each θ_i^{t-1} is the a posteriori expectation of $\theta_i^{t-1}|D_t$

$$\hat{\theta}_i^{t-1} = E\left[\theta_i^{t-1}|D_t\right]$$

In order to study the goodness of the proposed model, we model $\{X_t\}$ as an homogeneous Markov chain with initial distribution over the ten possible states

$$\pi_0 = [0, 0, 0, 0, 0, 0, 0, 0.3, 0.4, 0.3]$$

and matrix of transition probabilities

$$P = \begin{bmatrix} 1 & & & & & & & & & \\ 0.4 & 0.6 & & & & & & & & \\ & 0.3 & 0.7 & & & & & & & \\ & & 0.4 & 0.6 & & & & & & \\ & & & 0.3 & 0.7 & & & & & \\ & & & & 0.2 & 0.8 & & & & \\ & & & & & 0.2 & 0.8 & & & \\ & & & & & & 0.1 & 0.9 & & \\ & & & & & & & 0.1 & 0.9 & \\ & & & & & & & & 0.4 & 0.6 \end{bmatrix}$$

Using this process 400 paths were simulated. These paths were observed in 46 consecutive time slices. Each of these paths was perturbated by a stochastic matrix $Q = (q_{ij})$. Thus, if $X_t = m$ then $Y_t = k$, $1 \leq k \leq 10$, is obtained with probability q_{mk}. These perturbated paths are considered as the observable variables $\{Y_t\}$ to be used in inference and forecasting. The algorithm has been implemented in C language.

For each time slice, the implemented algorithm provides the a posteriori probability distribution $P(X_t|D_t)$. On the one hand, we compare the theoretical distribution of the real state of the equipment (given by $\pi_t = \pi_0 P^n$) in a set of time slices (specifically, in $t = 5$ and $t = 45$, which correspond with the system behaviour in a short and long-term) with the average bar diagram obtained by the algorithm over the 400 paths in those time slices. On the other hand, we study the solution provided to a particular path, summarizing each a posteriori distribution in its mean. Since we obtained similar results both in inference and forecasting, we only show up inference results.

First of all, we consider the identity matrix as the perturbation matrix, i.e., $Q = I_{10}$. In this way, the model receives the true system state as its observation and we try to study the goodness of the algorithm in ideal conditions. Figures 8, 9 and 10 show up the results. We observe that the model correctly identifies the real behaviour of the system, even in short-term. Dashed lines in Figure 10 stand for a 95% probability band.

Next, a sensibility study is proposed, using the perturbation matrix

$$Q_2 = \begin{bmatrix} 0.5 & 0.3 & 0.2 & & & & & & & \\ 0.2 & 0.5 & 0.2 & 0.1 & & & & & & \\ 0.1 & 0.2 & 0.4 & 0.2 & 0.1 & & & & & \\ & 0.1 & 0.2 & 0.4 & 0.2 & 0.1 & & & & \\ & & 0.1 & 0.2 & 0.4 & 0.2 & 0.1 & & & \\ & & & 0.1 & 0.2 & 0.4 & 0.2 & 0.1 & & \\ & & & & 0.1 & 0.2 & 0.4 & 0.2 & 0.1 & \\ & & & & & 0.1 & 0.2 & 0.4 & 0.2 & 0.1 \\ & & & & & & 0.1 & 0.2 & 0.5 & 0.2 \\ & & & & & & & 0.2 & 0.3 & 0.5 \end{bmatrix}$$

In short-term (Figure 11), there is a difference in the probabilities. However, an effective filtration of the considered perturbation is observed in long-term (Figure 12). In any case, Figure 13 shows the closeness of the inferred path to the real path in periods of moderate perturbation.

6 CONCLUSIONS

In our paper, we propose an approximate simulation algorithm for making inferences and forecasts about the states of non observable variables in partially observed dynamic systems modelled by chain graphs.

We want to emphasize that beliefs about the states at time t are adjusted from beliefs at time $t-1$ by the evidence collected at time t and included in the probability $P_{0t}(H_t|H_{t-1} = h_{t-1}, E_t = e_t)$. For instance, [Kanazawa et al.,95] use a similar argument in their *Evidence reversal* algorithm for dynamic probabilistic networks.

Our model is different from the standard hidden Markov model (used, for instance, in [Kanazawa et al.,95]) due to the kind of relationships fixed between observable and non observable variables. It seems to require a greater effort in the specification of probabilities. However, it allows to simplify substantially the exact and approximate procedures of inference and forecasting.

This greater effort is arguable, specially when cause-effect relationships cannot be assumed. In the example, we think it is easier to specify $P_t(\theta_i^{t-1}|\theta_i^{t-2}, Y_t)$ than $P_t(\theta_i^{t-1}|\theta_i^{t-2})$ (required when using a hidden Markov model). The first probability fixes our beliefs about the parameters based *simultaneously* on the previous state of these parameters *and* the evidence received.

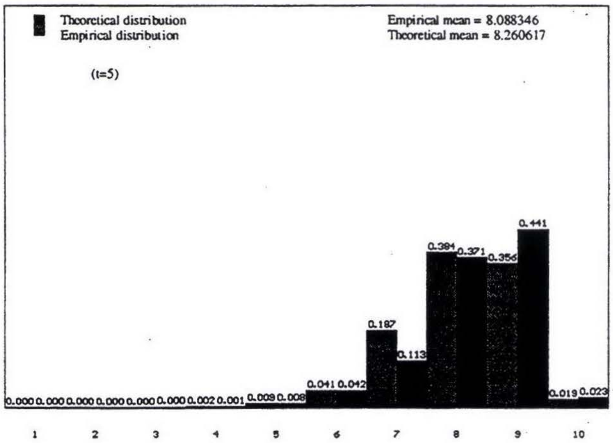

Figure 8

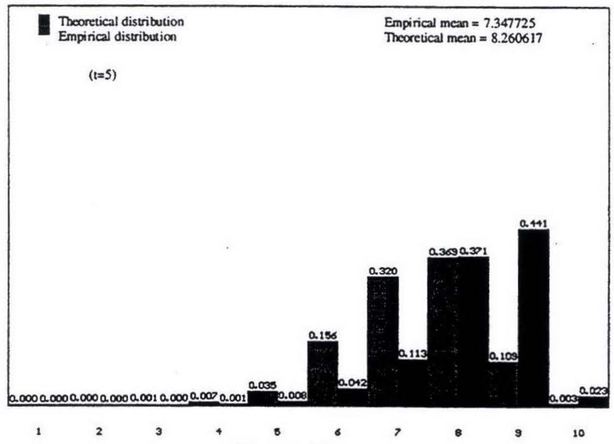

Figure 11

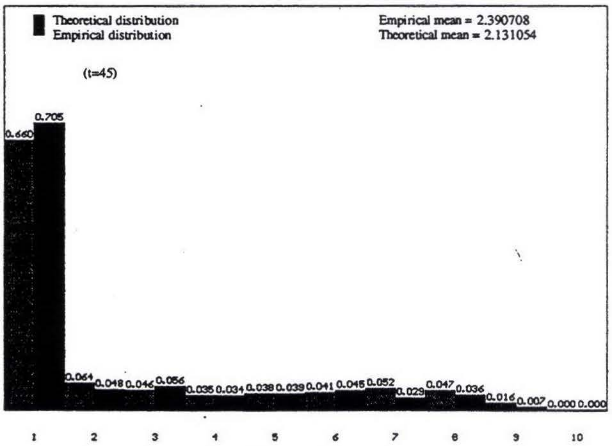

Figure 9

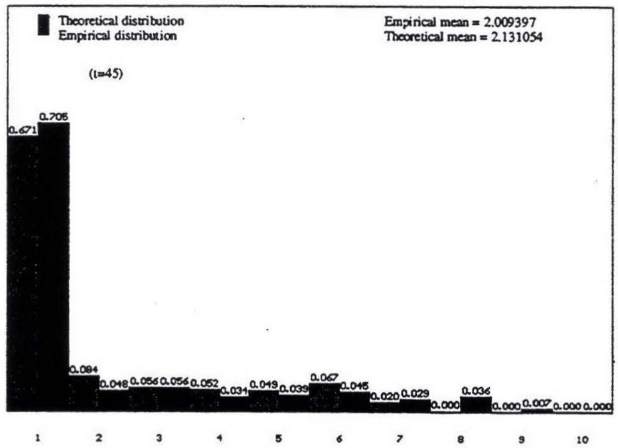

Figure 12

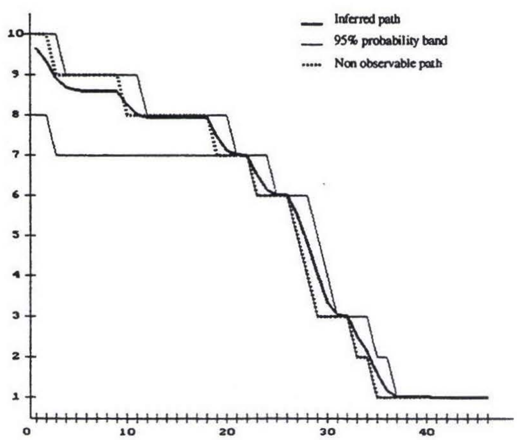

Figure 10

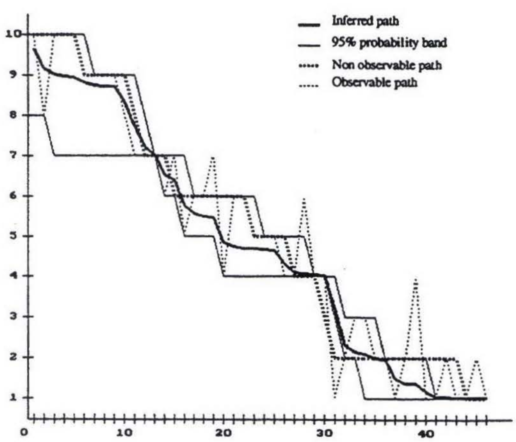

Figure 13

REFERENCES

[Berzuini,90] Berzuini, C. Modelling temporal processes via belief networks and Petri nets, with applications to expert systems. *Annals of Mathematics and Artificial Intelligence*, 2, 39-64, 1990.

[Dagum,Galper,94] Dagum, P.; Galper, A. *Time-series prediction using belief network models.* Technical Report 522, Knowledge Systems Laboratory, Medical Computer Science, Standford University. April 1994.

[Hanks et al.,95] Hanks, S.; Madigan, D.; Gavrin, J. Probabilistic temporal reasoning with endogenous change. In *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*, Montreal, Quebec, Canada, 18-20 August, 1995.

[Kanazawa et al.,95] Kanazawa, K; Koller, D.; Russell, S. Stochastic simulation algorithms for dynamic probabilistic networks. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI-95)*, 1995.

[Kjærulff,92] Kjærulff, U. A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence (UAI-92)*, Stanford University, 17-19 July, 121-129, 1992.

[Lekuona et al.,96] Lekuona, A.; Lacruz, B.; Lasala, P. Modeling and monitoring dynamic systems by chain graphs. In *Learning from Data: Artificial Intelligence and Statistics V.* Lecture Notes in Statistics, 112. D. Fisher and H.J. Lenz, eds., 68-77, NY, Springer-Verlag, 1996.

[Provan,94] Provan, G. Model selection for diagnosis and treatment using temporal influence diagrams. In P. Cheeseman y R.W. Oldford (eds.) *Selecting Models from Data*, Springer-Verlag, 133-142, 1994.

[Silver,63] Silver, E.A. *Markovian decision processes with uncertain transition probabilities or rewards.* Technical Report 1, Operation Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1963.