Representing Hypoexponential Distributions in Continuous Time Bayesian Networks

Manxia Liu[1,4(✉)], Fabio Stella[5], Arjen Hommersom[1,2], and Peter J. F. Lucas[1,3]

[1] Radboud University, ICIS, Nijmegen, The Netherlands
{m.liu,arjenh,peterl}@cs.ru.nl
[2] Open University of the Netherlands, Heerlen, The Netherlands
[3] Leiden University, LIACS, Leiden, The Netherlands
[4] University of Porto, CINTESIS, Porto, Portugal
[5] University of Milano-Bicocca, Milan, Italy
stella@disco.unimib.it

Abstract. Continuous time Bayesian networks offer a compact representation for modeling structured stochastic processes that evolve over continuous time. In these models, the time duration that a variable stays in a state until a transition occurs is assumed to be exponentially distributed. In real-world scenarios, however, this assumption is rarely satisfied, in particular when describing more complex temporal processes. To relax this assumption, we propose an extension to support the modeling of the transitioning time as a hypoexponential distribution by introducing an additional hidden variable. Using such an approach, we also allow CTBNs to obtain *memory*, which is lacking in standard CTBNs. The parameter estimation in the proposed models is transformed into a learning task in their equivalent Markovian models.

Keywords: Continuous time Bayesian networks
Dynamic Bayesian networks · Phase-type distribution · Memory
Hidden variable

1 Introduction

Continuous time Bayesian networks, or CTBNs for short, firstly introduced by Nodelman et al. [1], offer a compact representation for modeling structured stochastic processes that evolve over continuous time. By providing an explicit representation of time, i.e., time acts as a continuous parameter, these models

have the advantage of representing a probability distribution over observations that are made at irregularly spaced points in time. The powerful expressiveness of CTBNs to model data in such a form has been demonstrated by numerous early work (see e.g. reliability modeling [2], network intrusion detection [3,4], heart failure modeling [5,6], and gene network construction [7]).

In spite of supporting time irregularity, CTBNs suffer from an important limitation in their expressiveness: the time that a variable stays in a state until transition follows an exponential distribution. This distribution occurs naturally when describing a process where events occur continuously and independently at a constant average rate. In real-world scenarios, however, the assumption is rarely satisfied. In particular, it is inappropriate to describe more complex temporal processes, such as business processes that model interpurchase times [8]. The limitation was firstly described by Nodelman and Horvitz [9]; subsequent work by Gopalratman et al. [10] focuses on Erlang-Coxian distributions to handle time duration. To overcome the limitation, two approaches were proposed by Nodelman et al. [11] to extend CTBNs to phase-type duration distributions, yielding a richer and more flexible distribution. The first approach is to add hidden *states* to the random variables of a CTBN, which is called the *direct approach*. Alternatively, a second and more elegant approach is to add hidden *variables* to a CTBN. From a practical point of view, this approach is attractive, because existing CTBN inference algorithms can be directly applied. For the direct representation, states have to be interpreted as a disjunction of hidden states, which is cumbersome and computationally expensive when using existing software packages. However, the question of how to add the hidden variables to the network structure and what constraints should be imposed on the structure of their parameters was left unresolved [11].

In this paper, we show that the hidden variable approach can be used to represent a large class of duration distributions described by hypoexponential distributions. These distributions significantly generalize the existing exponential distributions. As a second contribution of this paper, we give precise conditions on the CTBN graph and discuss the exact constraints on the parameter structure for representing these distributions. We also show how these models are formally related to the direct representation, which we use for learning the parameters of the model.

The rest of the paper is organized as follows. We start with a motivating example in Sect. 2, followed by a brief summary of CTBNs and hypoexponential distribution in Sect. 3. Then, in Sect. 4, we define hidden continuous time Bayesian networks (HCTBNs). In Sect. 5, we show the relationships between the hidden variable model and the direct models. Subsequently, in Sect. 6, we demonstrate the usefulness of our proposed models by describing non-exponential distribution using HCTBNs and CTBNs, and by modeling dynamics of a medical problem. Finally, the paper is concluded with a brief discussion of possible future work for HCTBNs.

2   Motivating Example

To illustrate the proposed theory, we consider a medical example, viz. factors that influence the cardiac output, i.e., the blood flow to and from the heart. According to the literature, the heart rate, defined in terms of the number of heart beats per minute, has a positive influence on the cardiac output. However, a reduced blood supply, thus oxygen supply, as the result of coronary artery disease, may give rise to a heart attack (myocardial infarction). Consequently, some of the heart muscle fibers will die and the heart may fail to comply with respect to its function as a pump, thus cardiac output will be negatively affected. With regard to the prognosis, (increased) heart rate may be considered a risk factor for myocardial infarction (this is the rationale behind treatment of coronary artery disease patients with beta-blocking drugs, such as propranolol, that decrease heart rate). This causal knowledge is formalized as a directed graph in Fig. 1. Diagnosis of a myocardial infarction is done by examining the shape of the ECG and by determining the levels of troponin (a protein that is released from the dying heart cells) in the blood. In the model we take into account that lab facilities (to determine an ECG and troponin levels in the blood) are not available, as is common in some developing countries. Thus, diagnosing a myocardial infarction in the common way is not an option. As a result, the observations solely consist of the heart rate. With respect to modeling, this also implies hidden causes, such as myocardial infarction, must be taken into consideration when assessing potential causes for reduced cardiac output. More importantly, remembering having a myocardial infarction in the past, which is called *memory*, can alter the evolution of cardiac output in the future. In the remainder of this paper, we propose a method to deal with modeling such hidden causes, in particular to describe the memory behavior of temporal processes that evolve continuously over time.
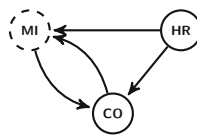

Fig. 1. Causal model for cardiac output: MI = Myocardial infarction; CO = cardiac output, HR = heart rate. The dashed node indicates a hidden cause.

3   Preliminaries

In this section, we will introduce the technical background of continuous time Bayesian networks as originally presented by Nodelman et al. [12] and phase-type distributions. The domain for an n-valued variable X is denoted as $\mathrm{Val}(X) = \{1, 2, \ldots, n\}$ with the notation $X = i$ indicating that variable X has the value i. We also use the notation $\pi(X)$ for the parents of variable X in a given graph.

3.1   Continuous Time Bayesian Networks

A *continuous-time Bayesian network* represents a stochastic process over a structured state space consisting of assignments to a set of local variables. The dynamics of the temporal evolution of the structured state space is described in terms of the evolution of the local variables. Let X be such a *local variable* with finite *domain* $\text{Val}(X) = \{1, 2, \ldots, n\}$, where $i \in \text{Val}(X)$ is called a *state*, and state changes over continuous time. The dynamics of X can be described as a homogeneous Markov process via its *intensity matrix*:

$$Q_X = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{pmatrix}$$

where $q_{ii} = -\sum_{j \neq i} q_{ij}$. The time that variable X stays in state i is exponentially distributed with rate $-q_{ii}$ and the expected time is given by $-1/q_{ii}$, and once it transitions from state i, it shifts to state j with probability $-q_{ij}/q_{ii}$.

CTBNs are based on homogeneous continuous time Markov processes, which has the Markov property, also known as a stronger assumption of *memorylessness*. The Markov property states that given the state of the process X at any set of times prior to time t, the distribution of X at time t depends *only* on X at the most recent time prior to time t. It is equivalent to say that given the state of the process X at time s, the distribution of X at any time after s is independent of the entire past of X prior to time s. More formally:

$$P(X_t = j \mid X_{t_1} = k_1, X_{t_2} = k_2, \ldots, X_{t_n} = k_n, X_s = i) = P(X_t = j \mid X_s = i)$$

where $0 < t_1 < t_2 < \cdots < t_n < s < t$.

However, this property has to be interpreted more carefully in CTBNs as these models also express the local dependence of one variable on the others. It is true that the Markov property still holds for CTBNs when conditioned on *all* the local variables in a model; it is not the case when only conditioning on a proper subset of the variables. This is due to the temporal entanglement in CTBNs where time is also considered. Let $\mathbf{X}$ be the variables in a CTBN and $\mathbf{Z}$ be a proper subset of $\mathbf{X}$, i.e., $\mathbf{Z} \subsetneq \mathbf{X}$. When querying the distribution over variables $\mathbf{Z}$ at time t, the distribution over variables $\mathbf{Z}$ at time t is no longer independent from its states at time prior to s given the states of variables $\mathbf{Z}$ at time s. More formally:

$$P(\mathbf{Z}_t = j, \mid \mathbf{Z}_{t_1} = k_1, \mathbf{Z}_{t_2} = k_2, \ldots, \mathbf{Z}_{t_n} = k_n, \mathbf{Z}_s = i) \neq P(\mathbf{Z}_t = j \mid \mathbf{Z}_s = i)$$

This is because the information at time prior to s is propagated to the time t through variables $\mathbf{X} \setminus \mathbf{Z}$. In this paper, we refer such behavior of variables $\mathbf{Z}$ as *memory*. Without loss of generality, we introduce memory to all variables in a

given CTBN by adding an additional *hidden* variable. The states of non-hidden variables are dependent on the their states in the past as the hidden variable is always unobservable. In this paper, we restrict ourselves to studying such memory behavior in CTBNs.

3.2   Hypoexponential Distribution

A phase-type distribution is a distribution which describes the time until reaching the absorbing state of a continuous time Markov chain with n transient states and one absorbing state. A phase-type distribution represented by n transient states is said to have order n. This continuous time Markov chain can be described as a state transition diagram. The diagram is a convenient graphical representation in terms of the initial probabilities, i.e., the distribution over the transient states at $t = 0$, the transition rates between the transient states, and the exit rates, i.e., the probability of entering the absorbing state.

Exponential distributions are a special case of phase-type distributions, where the continuous time Markov process has one transient state. The distribution can thus be graphically represented by a state transition diagram with only one state as shown in Fig. 2a. The diagram asserts that the chain enters the first and only transient state 1 with probability one and enters the absorbing state with rate λ. The *hypoexponential distribution*, also known as *generalized Erlang distribution*, is the distribution of the sum of n independent and identically exponentially distributed random variables. The state transition diagram of the Markov chain of n-order hypoexponential distribution is shown in Fig. 2. More details about phase-type distribution can be found in [13].
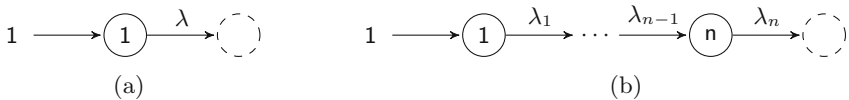

Fig. 2. State transition diagram for exponential distribution as shown in (a) and an n-order hypoexponential distribution as shown in (b). A solid node indicates a transient state and a dashed node indicates an absorbing state.

4   Hidden Continuous Time Bayesian Networks

In this section, we define a new extension of CTBNs, which we call hidden continuous time Bayesian networks, abbreviated to HCTBNs, where there is *only* one variable whose time staying in a state until a transition occurs, given a particular configuration of its parents, is described by a hypoexponential distribution. For other variables, the transition times are exponentially distributed.

4.1   Structure

First, we define the structure associated to an HCTBN, which contains a labeled node X which will correspond to a binary hypoexponential variable, a labeled node H corresponding to a hidden variable that is used to represent the hypoexponential distribution for X, and labeled $\mathbf{Y}$ to exponential variables.

Definition 1 (HCTBN Graph). *An HCTBN graph is a labeled graph defined by a triple $G = (\mathbf{V}, \mathbf{E}, l)$, where $\mathbf{V} = \{X, H\} \cup \mathbf{Y}$ denotes a set of vertices, $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$ a set of arcs on $\mathbf{V}$, and l a label function such that $l(X) = hypoexponential$, $l(H) = hidden$ and $l(\mathbf{Y}) = exponential$. The following conditions apply to G:*

1. *$H \to X \in \mathbf{E}$ and $X \to H \in \mathbf{E}$;*
2. *For any vertex $Y \in \mathbf{Y}$, $Y \to X \in \mathbf{E}$ iff $Y \to H \in \mathbf{E}$;*
3. *For any vertex $Y \in \mathbf{Y}$, $H \to Y \notin \mathbf{E}$.*

Condition 1 asserts that there is a bidirected edge between vertices X and H. Second, Condition 2 asserts as a parent of vertex X, vertex Y is also a parent of the hidden variable H. Together with Condition 1, it is clear that vertices X and H have the same number of parents. Thus, the number of parameters for H grows exponentially with the number of the parents of vertex X. Third, Condition 1 and 3 state that vertex X is the only child for vertex H.

Example 1. Consider the two simplest HCTBN graphs where we have two vertices X and H. In the first case, we have no other vertices, i.e., $\mathbf{Y} = \emptyset$. In the second case, we have another vertex Y and it is a parent of vertex X, i.e., $\mathbf{Y} = \{Y\}$. The HCTBNs graphs are given in Fig. 3.

Fig. 3. Two simplest HCTBNs graphs where vertex X has no children: (a) $\mathbf{Y} = \emptyset$ and $\pi(X) = \{H\}$; (b) $\mathbf{Y} = \{Y\}$ and $\pi(X) = \{H, Y\}$.

4.2   Model Definition

Now we give a formal definition of HCTBNs.

Definition 2 (Hidden   Continuous   Time   Bayesian   Networks (HCTBNs)). *An n-order hidden continuous time Bayesian network (HCTBN) is a triple $\mathcal{N} = (G, \Lambda, P_0)$ with the graph G as defined in Definition 1. In addition,*

Λ is a set of conditional intensity matrices and P_0 is the initial distribution for the variables associated to the nodes in the graph G with $P_0(X = 1, H = 1) = 1$, and for each configuration $\mathbf{u}$ of the parents $\mathbf{U}$ for variable X, $\mathbf{U} = \pi(X) \setminus \{H\}$, the intensity matrices for variable X and H have the following form:

$$Q_{H|X=1,\mathbf{u}} = \begin{pmatrix} -\lambda_1^{\mathbf{u}} & \lambda_1^{\mathbf{u}} & \cdots & 0 & 0 \\ 0 & -\lambda_2^{\mathbf{u}} & \lambda_2^{\mathbf{u}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & -\lambda_{n-1}^{\mathbf{u}} & \lambda_{n-1}^{\mathbf{u}} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad Q_{H|X=2,\mathbf{u}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ \gamma_{n-1}^{\mathbf{u}} & -\gamma_{n-1}^{\mathbf{u}} & \cdots & 0 & 0 \\ 0 & \gamma_{n-2}^{\mathbf{u}} & -\gamma_{n-2}^{\mathbf{u}} & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \gamma_1^{\mathbf{u}} & -\gamma_1^{\mathbf{u}} \end{pmatrix}$$

$$Q_{X|H=1,\mathbf{u}} = \begin{pmatrix} 0 & 0 \\ \gamma_n^{\mathbf{u}} & -\gamma_n^{\mathbf{u}} \end{pmatrix} \quad Q_{X|H=n,\mathbf{u}} = \begin{pmatrix} -\lambda_n^{\mathbf{u}} & \lambda_n^{\mathbf{u}} \\ 0 & 0 \end{pmatrix} \text{ If } n \geq 3, \quad Q_{X|H=2:n,\mathbf{u}} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

The intensity matrices defined in such a form make sure that the time duration distribution for variable X staying in a state is represented by a Markov chain with n transient states.

Example 2. Given the graph where $\mathbf{U} = \emptyset$ as shown in Fig. 3a, and $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3, \gamma_1 = 4, \gamma_2 = 5, \gamma_3 = 6$, we can define a 3-order HCTBN by giving the intensity matrices for variable X and H as below:

$$Q_{H|X=1} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \end{pmatrix} \quad Q_{H|X=2} = \begin{pmatrix} 0 & 0 & 0 \\ 5 & -5 & 0 \\ 0 & 4 & -4 \end{pmatrix}$$

$$Q_{X|H=1} = \begin{pmatrix} 0 & 0 \\ 6 & -6 \end{pmatrix} \quad Q_{X|H=2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad Q_{X|H=3} = \begin{pmatrix} -3 & 3 \\ 0 & 0 \end{pmatrix}$$

Alternatively, we can view the hypoexponential variable X and the hidden variable H as a whole by amalgamating them into a single variable S, whose state space is the joint state space over X and H. Each state of X now corresponds to a set of instantiations to S. When we amalgamate over the hypoexponential variable X and the hidden variable H, their joint intensity matrix follows a particular structure. The states in the intensity matrix are given by iterating over all the values of X in the ordering before iterating to the next values of H. In this particular case, this gives:

	11	21	12	22	13	23	$\cdots$	$1n-2$	$2n-2$	$1n-1$	$2n-1$	$1n$	$2n$	
	$-\lambda_1$	0	λ_1	0	0	0	0	0	0	0	0	0	0	11
	γ_n	$-\gamma_n$	0	0	0	0	0	0	0	0	0	0	0	21
	0	0	$-\lambda_2$	0	λ_2	0	0	0	0	0	0	0	0	12
	0	γ_{n-1}	0	$-\gamma_{n-1}$	0	0	0	0	0	0	0	0	0	22
$Q_{\mathrm{XH}} =$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots \cdots \cdots \cdots$		$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$		$\cdots$
	0	0	0	0	0	0	0	0	0	$-\lambda_{n-1}$	0	λ_{n-1}	0	$1n-1$
	0	0	0	0	0	0	0	0	$\gamma 2$	0	$-\gamma_2$	0	0	$2n-1$
	0	0	0	0	0	0	0	0	0	0	0	$-\lambda_n$	λ_n	$1n$
	0	0	0	0	0	0	0	0	0	0	γ_1	0	$-\gamma_1$	$2n$

Example 3. Consider the HCTBN as given in Example 2. The joint intensity matrix for variable X and H is given as below:

$$Q_{\mathrm{XH}} = \begin{array}{c} \begin{array}{cccccc} 11 & 21 & 12 & 22 & 13 & 23 \end{array} \\ \left(\begin{array}{cccccc} -1 & 0 & 1 & 0 & 0 & 0 \\ 6 & -6 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 2 & 0 \\ 0 & 5 & 0 & -5 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3 & 3 \\ 0 & 0 & 0 & 4 & 0 & -4 \end{array} \right) \begin{array}{c} 11 \\ 21 \\ 12 \\ 22 \\ 13 \\ 23 \end{array} \end{array}$$

Proposition 1. *Let X be the hypoexponential variable in an n-order HCTBN. The time that variable X stays in each of its states follows an n-order hypoexponential distribution.*

Given the joint intensity matrix of the hypoexponential variable X and the hidden variable H in an n-order HCTBN, now we can reinterpret the time duration of variable X in terms of the joint state over variables X and H. More specifically, the time of variable X staying in a state is then reinterpreted as the absorbing time of a Markov chain with a sequence of joint states over variable H and X where variable X in the joint states remains in the given state. For example, the time of variable X stays in state 1 is thus viewed as the absorbing time of a Markov chain with a sequence of joint states $11, 12, \ldots, 1n$, where X always stays in state 1 and the final transition in such a chain is the transition from state $1n$ to $2n$. As noted, there is a no explicit absorbing state. It is clear that such a Markov chain describes an n-order hypoexponential distribution. Analogously, we can construct another Markov chain corresponding to state 2 for variable X. Together, we can obtain a single Markov chain that could be graphically represented by a cyclic state transition diagram as shown in Fig. 4a.
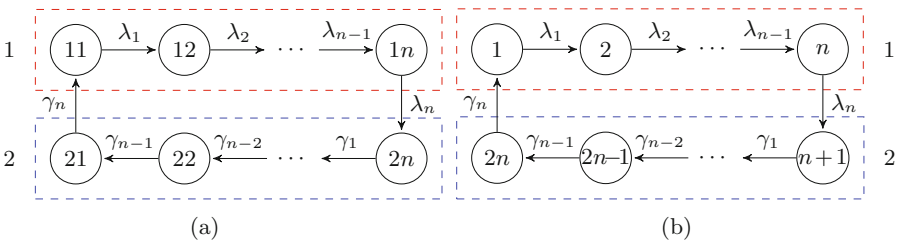

Fig. 4. State transition diagram for joint states over the hypoexponential variable X and hidden variable H in an n-order HCTBN (a) and for states of its extended variable X' in its equivalent Markovian model (b).

5   Equivalent Markovian Models

An important task for any probabilistic graphical models is to estimate parameters from data. In this paper, we transform parameter estimation in HCTBNs

into a learning task in their equivalent Markovian models in terms of the same time distribution for the hypoexponential variable. In this section, we devote ourselves to defining such equivalent Markovian models. The introduction of these models only serves as a tool to estimate parameters in HCTBNs.

Definition 3 (Equivalent Markovian Graph). *Let $G = (\mathbf{V}, \mathbf{E}, l)$, $\mathbf{V} = \{X, H\} \cup \mathbf{Y}$ be an HCTBN graph. An* equivalent Markovian graph $G' = (\mathbf{V}', \mathbf{E}')$ *is obtained with vertices $\mathbf{V}' = \{X\} \cup \mathbf{Y}$ and arcs $\mathbf{E}' = \mathbf{E} \cap (\mathbf{V}' \times \mathbf{V}')$.*

Hence, the graph structure is restricted by excluding the hidden variable H in the graph G' while all other variables remain. However, a different distribution is associated to vertex X in G', in particular the state-space has grown. For example, the equivalent Markovian model graphs associated to HCTBNs introduced in Fig. 3 are shown in Fig. 5.

(a)   (b)

Fig. 5. Equivalent Markovian graphs associated to HCTBNs as introduced in Fig. 3: (a) $\mathbf{Y} = \pi(X) = \emptyset$; (b) $\mathbf{Y} = \pi(X) = \{Y\}$.

Definition 4 (Equivalent Markovian Models). *Let $\mathcal{N}$ be an n-order HCTBN with intensity matrices Λ. An* equivalent Markovian model $\mathcal{M}$ *is defined as a triple $\mathcal{M} = (G', \Lambda', P_0')$ where graph $G' = (\{X\} \cup \mathbf{Y}, E)$ as in Definition 3, Λ' a set of intensity matrices over the vertices of G', and P_0' the initial distribution with $P_0'(X = 1) = 1$. For any $Y \in \mathbf{Y}$, if $X \notin \pi(Y)$, $Q_{Y|\pi(Y)}^{\mathcal{M}} = Q_{Y|\pi(Y)}^{\mathcal{N}}$; otherwise, $Q_{Y|\mathbf{K}, X=1:n}^{\mathcal{M}} = Q_{Y|\mathbf{K}, X=1}^{\mathcal{N}}$ and $Q_{Y|\mathbf{K}, X=n+1:2n}^{\mathcal{M}} = Q_{Y|\mathbf{K}, X=2}^{\mathcal{N}}$, where $\mathbf{K} = \pi(Y) \setminus \{X\}$. Given each configuration $\mathbf{u}$ of parents $\pi(X)$ from $\mathcal{M}$ and joint intensity matrix $Q_{XH}^{\mathcal{N}}$, intensity matrices $Q_{X|\pi(X)=\mathbf{u}}^{\mathcal{M}}$ are defined by re-ordering the states of $Q_{XH}^{\mathcal{N}}$ from current indices $[1, \ldots, 2n]$ to $[1, 3, \ldots, 2n-1, 2n, \ldots, 4, 2]$.*

Definition 4 implies that HCTBNs have the same number of parameters in their equivalent Markovian models.

Example 4. An equivalent Markov model for the HCTBN, as defined in Example 2, with the intensity matrix for variable X is given as below:

$$
Q_X = \begin{pmatrix}
 & 1 & 2 & 3 & 4 & 5 & 6 & \\
-1 & 1 & 0 & 0 & 0 & 0 & & 1 \\
0 & -2 & 2 & 0 & 0 & 0 & & 2 \\
0 & 0 & -3 & 3 & 0 & 0 & & 3 \\
0 & 0 & 0 & -4 & 4 & 0 & & 4 \\
0 & 0 & 0 & 0 & -5 & 5 & & 5 \\
6 & 0 & 0 & 0 & 0 & -6 & & 6
\end{pmatrix}
$$

Proposition 2. *Let X be the hypoexponential variable in an n-order HCTBN and X' be the extended variable of X in its associated equivalent Markov model. The absorbing time in a Markov chain described by a sequence of states $1, 2, \ldots, n$ of variable X' follows the same distribution as the time distribution of X staying in state 1, and the absorbing time in a Markov chain described by a sequence of states $n + 1, n + 2, \ldots, 2n$ of variable X' follows the same distribution as the time distribution of X staying in state 2.*

Similar to an HCTBN, we can also construct a state transition diagram for its equivalent Markovian model, as shown in Fig. 4b. The time that X staying in state 1 has an n-order hypoexponential distribution with rates $\lambda_1, \lambda_2, \ldots, \lambda_n$. The same distribution can also be represented by a Markov chain of a sequence of states of variable X', $1, 2, \ldots, n$. The same applies to X staying in state 2.

6   Experiments

In the experiments, we investigate two aspects. First, we investigate whether HCTBNs provide a better approximation than CTBNs when the true temporal processes are governed by a hypoexponential time distribution. Second, we show the usefulness of HCTBNs by modeling a number of factors that influence cardiac output in the medical setting, which was previously introduced in Sect. 2. In this model, an interesting question is how the dynamics of cardiac output are affected by other factors, in particular when a hidden cause is present, i.e., when myocardial infraction is not observed.
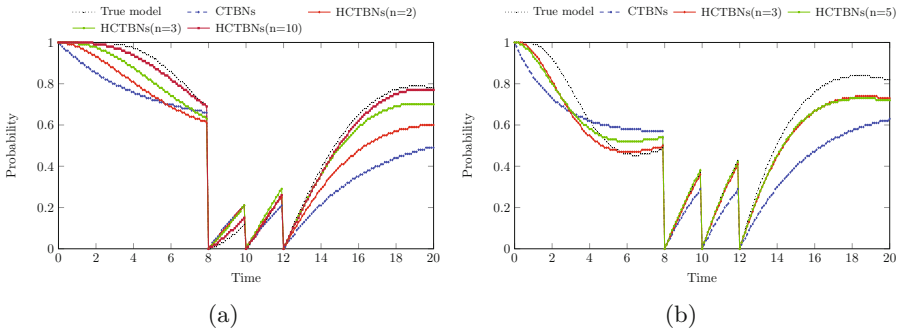

Fig. 6. Probability of X staying at state 1, given evidence $X = 2$ and $Y = 2$ at time 8, 10 and 12. (a): the true process has 10-order hypoexponential distribution and no parents. (b): the true process has 5-order hypoexponential distribution and one parent. The rates in the distribution follow a *Gamma distribution* with rate $= 1$ and shape $= 2$. The number of hidden states for the learned HCTBNs is indicated by the number n.

In the experiments, two software packages were mainly used to learn parameters for HCTBNs. The transformation between a given HCTBN and its

equivalent Markovian model was implemented by *CTBN-RLE*[1]. We reformulated the parameter estimation task in HTCBNs as one in their equivalent Markovian models, where the EM algorithm is used to approximate a phase-type distribution from data by employing *EMpht*[2]. The EMpht also supports learning parameters from *right censored data*, i.e., a variable staying in a state for at least a given amount of time. A more detailed discussion can be found in [10].

For the first purpose, we generated a number of datasets from temporal processes where the time distribution follows a more complex distribution, rather than simple exponential distribution. In the experiments, a hypoexponential distribution was chosen. With respect to learning parameters for HCTBNs, we also considered the impact of the number of hidden states on the quality of the approximation in learned HCTBNs. The number of hidden states was set to 2, 3 and 10 when the underlying hypoexponential distribution has an order 10, and to 3 and 5 when the distribution has order 5.

For illustrative purpose, we considered learning parameters for a variable with complex time distribution without parents as shown in Fig. 5a and in the presence of one single parent as shown in Fig. 5b. The underlying time distribution was approximated by using the proposed HCTBNs and CTBNs. The dynamics of the hypoexponential variable X in the time interval [0, 20] in learned CTBNs and HCTBNs, as shown in Fig. 6, suggest that HCTBNs have a better approximation of the underlying generalized hypoexponential distribution than CTBNs. It also indicates that other complex distributions may be better approximated using HCTBNs. In addition, we obtained a better approximation using HCTBNs with more hidden states. More importantly, the memory in a given temporal process can be easily captured by HCTBNs, whereas it can not be captured using CTBNs.
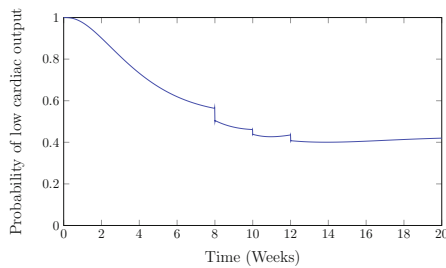

Fig. 7. Probability of having a low cardiac output given the evidence of high heart rate at time 8, 10 and 12.

For the second part of the experiments, we show the usefulness of HCTBNs for the medical example by modeling the dynamics of a patient's cardiac output over time. We computed the probability distribution of cardiac distribution for

---

a period of 20 weeks. At time 0, the patient has a myocardial infarction and evidence of the patient having high heart rate is given at time 8, 10, 12. Results of this experiment are plotted in Fig. 7. The plot shows that it is less likely for the patient to have a low cardiac output given a high heart rate (see drops at time 8, 10 and 12). The plot also suggests that factors that influence cardiac output cannot be solely explained by heart rate as we have different probabilities of having a low cardiac output at time 8, 10 and 12, even given the same evidence.

7   Conclusions

In this paper, we show that time duration in CTBNs governed by hypoexponential distributions can be modeled by using hidden variables. In addition, we show that the hidden variable also introduces *memory*, which is lacking in standard CTBNs. This memory will make CTBNs better-suited as a modeling tool for more general real-world problems in many domains, such as biology where memory plays a central role. In this paper, we provide a complete formalization of the approach. In addition, experimental results show that HCTBNs indeed can learn this more complex distributions, which was also illustrated by a small medical example.

A limitation of HCTBNs so far is that the observable variables are restricted to two states, as the focus of this paper has been on the introducing a richer time distribution and memory. In future work, we aim to overcome this limitation by supporting multinomial variables. At first glance, the proposed procedure for transforming to equivalent Markovian models can also be applied for multinomial variables but a further careful examination is necessary.

References

1. Nodelman, U.D.: Continuous time Bayesian Networks. Ph.D. thesis, Stanford University (2007)
2. Boudali, H., Dugan, J.B.: A continuous-time Bayesian network reliability modeling, and analysis framework. IEEE Trans. Reliab. **55**(1), 86–97 (2006)
3. Xu, J., Shelton, C.R.: Continuous time Bayesian networks for host level network intrusion detection. In: Daelemans, W., Goethals, B., Morik, K. (eds.) ECML PKDD 2008. LNCS (LNAI), vol. 5212, pp. 613–627. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87481-2_40
4. Xu, J., Shelton, C.R.: Intrusion detection using continuous time Bayesian networks. J. Artif. Intell. Res. **39**, 745–774 (2010)
5. Gatti, E., Luciani, D., Stella, F.: A continuous time Bayesian network model for cardiogenic heart failure. Flex. Serv, Manuf. J. **24**(4), 496–515 (2012)
6. Liu, M., Hommersom, A., van der Heijden, M., Lucas, P.J.F.: Hybrid time Bayesian networks. Int. J. Approximate Reasoning **80**, 460–474 (2017)
7. Acerbi, E., Viganò, E., Poidinger, M., Mortellaro, A., Zelante, T., Stella, F.: Continuous time Bayesian networks identify Prdm1 as a negative regulator of TH17 cell differentiation in humans. Scientific reports 6 (2016)
8. Angus, I.: An introduction to Erlang B and Erlang C. Telemanagement **187**, 6–8 (2001)

9. Nodelman, U., Horvitz, E.: Continuous time Bayesian networks for inferring users' presence and activities with extensions for modeling and evaluation. Microsoft Research, July–August 2003

10. Gopalratnam, K., Kautz, H., Weld, D.S.: Extending continuous time Bayesian networks. In: Proceedings of the National Conference on Artificial Intelligence, vol. 20, p. 981. AAAI Press, MIT Press; 1999, Menlo Park, Cambridge, London (2005)

11. Nodelman, U., Shelton, C.R., Kollerthu, D.: Expectation maximization and complex duration distributions for continuous time Bayesian networks. In: Proceedings of the Twenty-First International Conference on Uncertainty in Artificial Intelligence, pp. 421–430 (2005)

12. Nodelman, U., Shelton, C.R., Koller, D.: Continuous time Bayesian networks. In: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, pp. 378–387 (2002)

13. Verbelen, R.: Phase-type distributions & mixtures of Erlangs (2013)