Making Continuous Time Bayesian Networks More Flexible

Manxia Liu LIUMANXIA@CS.RU.NL
Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands

Fabio Stella STELLA@DISCO.UNIMIB.IT
Department of Informatics, Systems and Communication, University of Milan-Bicocca, Italy

Arjen Hommersom ARJEN.HOMMERSOM@OU.NL
Faculty of Management, Science, and Technology, Open University of the Netherlands
Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands

Peter J. F. Lucas PETERL@CS.RU.NL
Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands
Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands

Abstract

The time duration in continuous time Bayesian networks, i.e., the time that a variable stays in a state until it transitions to another state, follows an exponential distribution. The exponential distribution is widely applied to describe the waiting time between events in a Poisson process, which describes the distribution of the number of events in one unit of time. This distribution is parameterized by a single rate and has mode zero, implying that the highest probability mass for events to happen is attributed to the earliest times. To describe biological processes, the exponential distribution is not always natural. For example, if the immune system has not encountered a pathogen before, it most likely responds to a viral infection after a few days, rather than immediately. In this paper, we generalize our recently proposed hypoexponential continuous time Bayesian networks, by allowing any number of hypoexponential variables, i.e., variables having a hypoexponential time duration distribution. In addition, we propose and compare two learning methods to estimate parameters for the generalized models. Finally, the practical value of the generalized models is demonstrated by means of a realistic medical problem.

Keywords: Continuous time Bayesian networks, Hypoexponential distribution, Parameter estimation.

1. Introduction

Describing waiting time, the time between events, is an important part of modeling real-world problems involving time. For example, a question of clinical interest concerning viral infections is how much time it takes for an individual to become infected after contact with an infected other individual. Waiting time in continuous time Bayesian networks (CTBNs) is described by an exponential distribution, a simple (with a single rate) but powerful distribution. However, to describe biological processes, the exponential distribution is not always natural. Clearly, an individual can not immediately get a viral infection or recover immediately from a viral disease.

A natural way to gain more flexibility is obtained by replacing the exponential distribution on time by a phase-type distribution. The phase-type distribution is described by a Markov chain over variables with exponential time distributions with the exponential distribution as a special case. In CTBNs, such a distribution can be represented by: (1) adding additional hidden states to a random variable, called the *direct approach*; (2) adding a hidden variable, called the *hidden approach*. The

direct method can explicitly give us the transitions between exponentially distributed variables in the Markov chain representing the phase-type distribution. A disadvantage of the direct approach is that we can only observe the current *state* of a variable, but not its associated hidden states. This imposes a challenge for using the direct approach to learn from data, as the hidden states are unknown in the data even when the data are complete. A trajectory in the complete data, i.e., when and where a transition occurs, can still be interpreted in a number of different ways using the hidden states. This makes it impossible to make use of existing learning methods for CTBNs. From an inference point of view, the query about whether a variable stays in a given state has to be represented by a disjunction of the hidden states, which is not always appropriate, in particular for describing interval evidence. This is because it requires more than one hidden state to represent phase-type distributions in general, except for the common exponential distribution. Interval evidence takes the form of a variable staying in a state during a given time interval. Although the state of the variable does not change in the time interval, there are myriad combinations of hidden states associated to the state of the variable to interpret the evidence. For example, we can choose an arbitrary hidden state or any arbitrary combinations of hidden states for the given time interval.

Using the hidden approach, a variable with phase-type time distribution represents a problem of interest in a given context, which is lacking in the direct approach. For example, in a medical context, a variable can represent the underlying medical condition, such as whether a patient has cancer or an infectious disease, which helps in interpreting the observed signs and symptoms. The use of a hidden variable also gives the hidden approach the potential of computational advantages in comparison to the direct approach.

In this paper, first our recently proposed hypoexponential continuous time Bayesian networks (Liu et al. (2018)), where there is only a single variable with hypoexponential time distribution—an example of a phase-type distribution—, is generalized by supporting any number of hypoexponential variables. Second, we propose and compare the direct and hidden approaches to estimate parameters from complete data for the generalized models. The direct approach is expected to be more computationally efficient than the hidden approach at the expense of the quality of the learned models. Third, the usefulness of the generalized framework is illustrated by a realistic medical problem where multiple hypoexponential variables are involved in the modeling.

2. Related Work

The suggestion to extend CTBNs by replacing the exponential distribution with particular phase-type distributions, such as Erlang and Coxian distributions, was first made in Nodelman and Horvitz (2003). It was followed by another suggestion of using two different approaches, the direct and hidden, to model phase-type time distributions in CTBNs (Nodelman et al. (2005)). As a phase-type distribution is a probability distribution constructed by an expression involving exponential distributions, not necessarily a weighted sum, there is some similarity to the use of the mixture of truncated exponentials (Moral et al. (2001)) that allows modeling any distribution, discrete or continuous. In a similar way, a phase-type distribution can provide an accurate approximation of any positive-valued distribution, and thus provides a greater variety of complex distributions, rather than the simple exponential distribution.

3. Motivating Example

To illustrate the usefulness of the proposed theory, we consider the medical problem of chronic obstructive pulmonary disease (COPD), with its causal graph given in Fig. 1. Obtaining insight into the evolution of a chronic disease is an important aspect of chronic disease management, as often the disease will not disappear. In the context of COPD, it is of particular importance to study the effect of a viral infection on lung function, as a viral infection is a major risk factor of a COPD exacerbation, i.e., the disease gets worse. Having a better understanding of the incubation time, i.e., the time for a COPD patient to get a viral infection, and the recovery time, i.e., the time that a COPD patient will recover from a worsening in lung function, is important for standardizing and optimizing the duration of medical treatment. As an example to illustrate the potential of the developed methods, we consider the modeling of incubation and recovery time of a viral infection of a COPD patient.

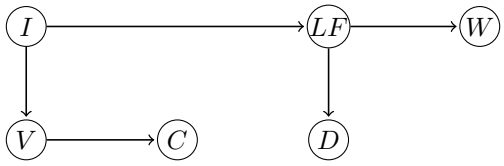

Fig. 1: Expert opinion based Bayesian network for COPD problem; I: infection, V: sputum volume, C: cough, LF: lung function, D: dyspnea, W: wheeze. Arcs stand for temporal dependences, same for all the other arcs in the reminder of this paper.

There have been many previous efforts to model the incubation time of viral infections. For example, according to Bailey (1954) and Gough (1977) the incubation time duration cannot be approximated by an exponential distribution, as the mode (the maximum value of the associated density) of the underlying distribution can be far away from zero, whereas the mode of the exponential distribution is actually zero. One example of such non-zero mode distributions is illustrated in Fig. 2, which is generated by simulating the incubation time distribution from empirical data from Bailey (1954); Gough (1977). It is obvious that such a distribution cannot be well-captured by the exponential distribution. Exponentially distributed events tend to occur close together, which is not an accurate model for the incubation time of viral infectious diseases. Incubation may take a certain amount of time, due to the fact that virus particles have to be replicated before they are present in sufficient quantity to affect the body.

For COPD patients, the lung function can deteriorate over time, characterized by worsening lung-related symptoms, such as dyspnea. In clinical practice, a peak flow meter is often used as a tool to measure lung function. In the literature, Seemungal et al. (2000) have shown that it takes one to two weeks for the majority of moderate to severe COPD patients to recover from their worsening lung function in terms of peak flow. This implies that the mode of the underlying time distribution for recovering from decreased lung function is non-zero.

The examples of incubation time and lung function recovery both indicate that the exponential distribution cannot offer a satisfactory representation for many disease processes. As the exponential distribution is a limitation of present CTBNs, we seek more versatile and flexible distributions to handle more complicated real-world problems.

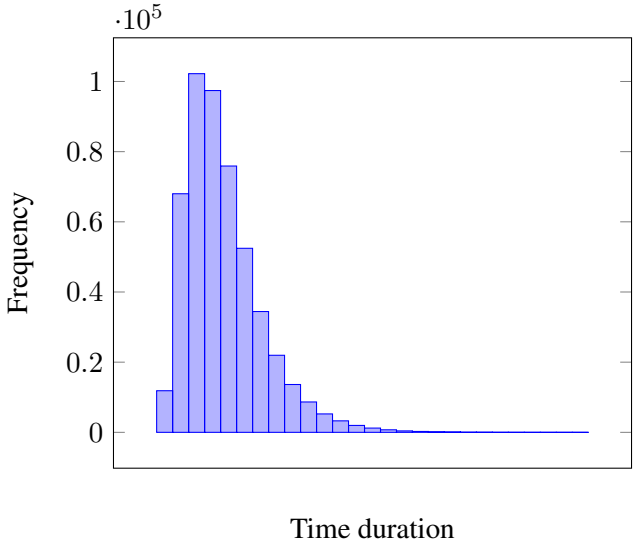

Fig. 2: A time distribution generated according to the incubation time in the work by Bailey (1954) and Gough (1977) based on non-zero mode.

4. Phase-type Distributions

A phase-type distribution is represented by a random variable describing the time that a finite-state absorbing continuous time Markov chain reaches its only absorbing state, i.e., a state that once entered cannot be left. Each state of the Markov chain represents a stochastic process with an exponential time distribution. Before the chain reaches its absorbing state, it moves through its transient states, i.e., a state that once it is reached, the probability of returning to the state is less than one after visiting it (Verbelen (2013)).

A CTBN is described by a continuous time Markov chain $\{X_t \mid t \in \mathbb{R}_0^+\}$ (the chain may be absorbing or non-absorbing) with n transient states and one absorbing state $n+1$, parameterized by an intensity matrix Q specifying the intensities for a state transitioning to another. To form intensity matrices for multiple processes, two intensity matrices are taken to perform an operation called *amalgamation* (Nodelman et al. (2002)). A phase-type distribution is described by an absorbing continuous time Markov chain with parameters fully specified by an initial distribution vector $\mathbf{p} = [p_1, \ldots, p_n]^T$ ($\mathbf{p}^T$ is the transpose of $\mathbf{p}$), $p_i = P(X_0 = i)$, $i \in \{1, \ldots, n\}$ over n transient states with $\sum_{i=1}^{n} p_i = 1$ (the probability for the absorbing state is zero). The intensity matrix Q for a phase-type distribution states that transitioning away from the absorbing state has zero intensity. More specifically, the intensity matrix Q is defined as follows:

$$Q = \begin{pmatrix} A & \mathbf{v} \\ \mathbf{0}^T & 0 \end{pmatrix}$$

where A is an $n \times n$ dimensional (square) matrix, specifying the intensities for transitioning between transient states, $\mathbf{v}$ is a column vector where v_i is the intensity leaving transient state i to the absorbing state, and $\mathbf{0}^T$ is a zero row vector of dimension n. The matrix A is called the phase-type generator, and n is called the order of the phase-type distribution. Every row in matrix Q sums to

zero, from which it follows that $\mathbf{v} = -A\mathbf{e}$, where $\mathbf{e} = [1, 1, \ldots, 1]^T$ is an n-dimensional column vector of ones.

The density function for the phase-type distribution is defined as:

$$f(t) = \mathbf{p}^T \exp(At) \cdot -A\mathbf{e} = \mathbf{p}^T \exp(At)\mathbf{v}$$

The distribution function for the phase-type distribution is:

$$F(t) \quad = \quad \int_0^t f(s)\mathrm{d}\, s = 1 - \mathbf{p}^T \exp(At)\mathbf{e}$$

For an n-order phase-type distribution, its associated continuous time Markov chain can be graphically represented by a state transition diagram. The diagram is a convenient graphical representation by specifying the initial probabilities $\mathbf{p}$, i.e., the distribution over the all the transient states when the chain starts, the rates between the transient states, and the exit rates, i.e., the rate for a transient state entering the absorbing state. More details about phase-type distributions can be found in Bladt (2005); Verbelen (2013).

In this paper, we mainly focus on the hypoexponential distribution, also known as *generalized Erlang distribution*. It is a rich and flexible subset of phase-type distributions. For a hypoexponential distribution, a transient state is only allowed to enter its consecutive transient state or the absorbing state in the corresponding Markov chain. In addition, it starts only in one of the transient states and traverses all the other transient states until it reaches the absorbing state. The n-order hypoexponential distribution is graphically represented by a Markov chain with its state diagram as shown in Fig. 3. The diagram asserts that the chain enters transient state 1 with probability one and enters the absorbing state with rate λ_n.

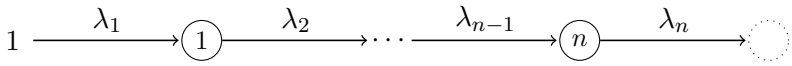

Fig. 3: A state transition diagram for an n-order hypoexponential distribution. A solid node indicates a transient state and a dashed node indicates an absorbing state. The number '1' at the left-hand side denotes the initial probability of entering a state.

The specification of the hypoexponential distribution consists of the vector of initial probabilities $\mathbf{p} = [1, 0, \ldots, 0]^T$, $\mathbf{v} = [0, 0, \ldots, \lambda_n]^T$ and the matrix A:

$$A = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & -\lambda_n \end{pmatrix}$$

5. Hidden Continuous Time Bayesian Networks

In this section, we generalize our recently proposed *hidden continuous time Bayesian networks* (Liu et al. (2018)), abbreviated to HCTBNs, where the time duration follows a hypoexponential distribution. In HCTBNs, variables are categorized into three groups, hypoexponentially distributed variables, auxiliary hidden variables, and exponentially distributed variables. The generalized models are defined in terms of the structure of a graph and parameters.

5.1 Structure

The graphical structure associated with an HCTBN, with labeled nodes $\mathbf{X}$ for binary hypoexponential variables, labeled nodes $\mathbf{H}$ for auxiliary hidden variables, and labeled nodes $\mathbf{Y}$ for exponential variables, is defined first. Values of a binary variable X come from $\{1, 2\}$; hidden and exponential variable take values from $\{1, \ldots, n\}$. The time duration of a hypoexponential variable follows a hypoexponential distribution, while it can only take two possible values, indicated by 1 and 2. The *parents* of a variable X in its corresponding graph are denoted by $\pi(X)$.

Definition 1 (HCTBN Graph) *An HCTBN graph is a node-labeled graph defined by a tuple $G = (\mathbf{V}, \mathbf{E}, b, l)$, where $\mathbf{V} = \mathbf{X} \cup \mathbf{H} \cup \mathbf{Y}$ denotes a set of nodes where $\mathbf{X}$, $\mathbf{H}$ and $\mathbf{Y}$ are mutually disjoint, $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$ a set of arcs on $\mathbf{V}$, b a bijective function $b : \mathbf{X} \to \mathbf{H}$, and l a label function such that $l(\mathbf{X}) =$ hypoexponential, $l(\mathbf{H}) =$ hidden, and $l(\mathbf{Y}) =$ exponential. Furthermore, the graph G has the following properties:*

1. *For any $X \in \mathbf{X}$, $b(X) \to X \in \mathbf{E}$ and $X \to b(X) \in \mathbf{E}$;*
2. *For any $Z \in \mathbf{Y} \cup \mathbf{X}$, $Z \to X \in \mathbf{E}$ iff $Z \to b(X) \in \mathbf{E}$, $X \in \mathbf{X}$;*
3. *For any $H \in \mathbf{H}$, $H \to Y \notin \mathbf{E}$;*
4. *For any $H, H' \in \mathbf{H}$, $H \to H' \notin \mathbf{E}$.*

Property 1 states that for any $X \in \mathbf{X}$, there is exactly one associated hidden variable H, $H = b(X)$. This also implies that in an HCTBN $\mathbf{X}$ and $\mathbf{H}$ have the same cardinality. In addition, there are three main restrictions on the arcs in the HCTBN graph. First, Property 1 asserts that node X is connected to its associated hidden node $b(X)$ by a bidirected arc. Second, Property 2 asserts that the associated hidden node $b(X)$ has an exponential node $Y, Y \in \mathbf{Y}$, as a parent if and only if Y is also a parent of X. Of course, Y can be a parent of more than one hypoexponential nodes, and thus be a parent of more than one hidden nodes. For example, we may have an exponential node Y be a parent of two distinct hypoexponential nodes X, X', i.e., $Y \to X$ and $Y \to X'$. According to Definition 1, node Y is also a parent of the associated hidden nodes $b(X)$ and $b(X')$, i.e., we also have $Y \to b(X)$ and $Y \to b(X')$ in the graph. The graph properties imply that for each hypoexponential node X, the node has the same number of parents as its associated hidden node $b(X)$. Thus, the number of parameters for the hidden node $b(X)$ grows exponentially with the number of parents of node X. Third, Properties 3-4 state that a hypoexponential node X is the only child for its associated hidden node $b(X)$. This implies that there are no direct connections between any two distinct hidden nodes.

Now we can describe the time distribution for infection I and lung function LF in the COPD network by modeling them as hypoexponential variables in an HCTBN.

Example 1 *For the COPD problem, the eight variables are categorized into three groups: hypoexponential variables $\mathbf{X} = \{I, LF\}$, exponential variables $\mathbf{Y} = \{V, C, D, W\}$ and hidden variables $\mathbf{H} = \{H_1, H_2\}$ with $b(I) = H_1$ and $b(LF) = H_2$. The corresponding HCTBN graph is depicted in Fig. 4. Hypoexponential nodes I and LF have hidden nodes H_1 and H_2 as parents, respectively. In addition, hypoexponential node LF has the hypoexponential variable I as a parent, thus its corresponding hidden node H_2 also has an incoming arc from node I. Furthermore, all the exponential nodes excluding node C are either a child of the hypoexponential node I or of LF. It is important to mention that hidden nodes H_1 and H_2 are not directly connected with the exponential variables.*

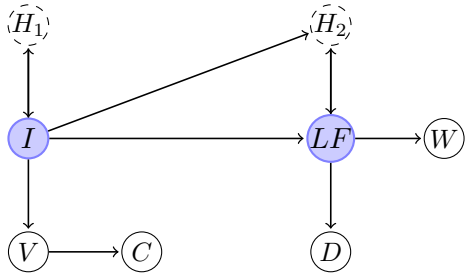

Fig. 4: An HCTBN graph for the COPD problem; I: infection, V: sputum volume, C: cough, LF: lung function, D: dyspnea, W: wheeze, and H_1 and H_2 are hidden nodes corresponding to I and LF, respectively. Dashed nodes correspond to hidden variables $\mathbf{H}$, solid blue nodes to hypoexponential variables $\mathbf{X}$, and solid nodes to the rest nodes $\mathbf{Y}$.

5.2 Model Definition

First, we generalize the HCTBN as originally defined in Liu et al. (2018).

Definition 2 (Hidden Continuous Time Bayesian Networks (HCTBNs)) *A hidden continuous time Bayesian network (HCTBN) is a triple $\mathcal{N} = (G, \mathbf{Q}, P_0)$ with the HCTBN graph G as defined in Definition 1. In addition, $\mathbf{Q}$ 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. For each $X \in \mathbf{X}$ with n_X-order hypoexponential distribution and $H = b(X)$, $n_X \in \mathbb{N}, n_X \geq 2$, we have either $P_0(X = 1, H = 1) = 1$ or $P_0(X = 2, H = n_X) = 1$. The following intensity matrices for variables X and H are defined: $Q_{H|X=1,\mathbf{U}=\mathbf{u}}$, $Q_{H|X=2,\mathbf{U}=\mathbf{u}}$, $Q_{X|H=1,\mathbf{U}=\mathbf{u}}$, $Q_{X|H=n_X,\mathbf{U}=\mathbf{u}}$, and for $n_X \geq 3$, $Q_{X|H\in\{2,\ldots,n_X-1\},\mathbf{U}=\mathbf{u}} = O$ (null matrix), where $\mathbf{U} = \pi(H) \cap \pi(X)$ and $\mathbf{u}$ are the assignments to the variables $\mathbf{U}$. Details are given in Liu et al. (2018).*

Note that the order for the hypoexponential distribution may differ from one hypoexponential variable to another, hence the use of the subscript X in n_X.

Example 2 *Consider an HCTBN with its HCTBN graph given in Fig. 4. In the model, we consider the intensity matrices for the hypoexponential variable LF and its corresponding hidden variable H_2. Suppose the hypoexponential variable LF has an n_{LF}-order hypoexponential distribution, where $n_{LF} = 4$, and we have parameters $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3, \lambda_4 = 4, \gamma_1 = 5, \gamma_2 = 6, \gamma_3 = 7, \gamma_4 = 8$ when $I = 1$, and $\lambda_1 = 9, \lambda_2 = 10, \lambda_3 = 11, \lambda_4 = 12, \gamma_1 = 13, \gamma_2 = 14, \gamma_3 = 15, \gamma_4 = 16$ when $I = 2$, this gives the intensity matrices for variable LF and H_2 shown in Fig. 5.*

In addition, hypoexponential variable LF has a number of exponential variables as children and we also consider the structure of the intensity matrix for one of its children D. There are no restrictions on the intensity matrices for variable D:

$$Q_{D|LF=1} = \begin{pmatrix} -29 & 29 \\ 30 & -30 \end{pmatrix} \quad Q_{D|LF=2} = \begin{pmatrix} -31 & 31 \\ 32 & -32 \end{pmatrix}$$

6. Parameter Learning

An important task for any probabilistic graphical models is to estimate parameters from data. As HCTBNs fit naturally into the CTBN framework, existing learning algorithms can be directly ap-

$$Q_{H_2|LF=1,I=1} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & -3 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$Q_{H_2|LF=2,I=1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 7 & -7 & 0 & 0 \\ 0 & 6 & -6 & 0 \\ 0 & 0 & 5 & -5 \end{pmatrix}$$

$$Q_{H_2|LF=1,I=2} = \begin{pmatrix} -9 & 9 & 0 & 0 \\ 0 & -10 & 10 & 0 \\ 0 & 0 & -11 & 11 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$Q_{H_2|LF=2,I=2} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 15 & -15 & 0 & 0 \\ 0 & 14 & -14 & 0 \\ 0 & 0 & 13 & -13 \end{pmatrix}$$

$$Q_{LF|H_2=1,I=1} = \begin{pmatrix} 0 & 0 \\ 8 & -8 \end{pmatrix}$$

$$Q_{LF|H_2\in\{2,3\},I=1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$Q_{LF|H_2=4,I=1} = \begin{pmatrix} -4 & 4 \\ 0 & 0 \end{pmatrix}$$

$$Q_{LF|H_2=1,I=2} = \begin{pmatrix} 0 & 0 \\ 16 & -16 \end{pmatrix}$$

$$Q_{LF|H_2\in\{2,3\},I=2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$Q_{LF|H_2=4,I=2} = \begin{pmatrix} -12 & 12 \\ 0 & 0 \end{pmatrix}$$

Fig. 5: Parameters for hypoexponential variable LF and its associated hidden variable H_2 in the HCTBN with its structure given in Fig. 4.

plied to estimate parameters for HCTBNs. Alternatively, HCTBNs can be transformed into their equivalent direct models having the same time distribution for hypoexponential variables. In this section, we define such equivalent direct models from given HCTBNs. The introduction of these models only serves as an alternative to estimate parameters for HCTBNs.

Definition 3 (Equivalent Direct Graph) *Let $G = (\mathbf{V}, \mathbf{E})$ be an HCTBN graph with hypoexponential nodes $\mathbf{X}$, hidden nodes $\mathbf{H}$ and exponential nodes $\mathbf{Y}$ and $\mathbf{V} = \mathbf{X} \cup \mathbf{H} \cup \mathbf{Y}$. An equivalent direct graph G' is defined as a graph $G' = (\mathbf{V}', \mathbf{E}')$, with nodes $\mathbf{V}' = \mathbf{X} \cup \mathbf{Y}$ and arcs $\mathbf{E}' = \mathbf{E} \cap (\mathbf{V}' \times \mathbf{V}')$.*

For an HCTBN graph G, an equivalent direct graph G' excludes all the hidden nodes $\mathbf{H}$ from the graph G, while it includes all the hypoexponential and exponential nodes. Any arcs linked to hidden nodes are also omitted in graph G'. For example, the hidden nodes H_1 and H_2 in the HCTBN in Fig. 4 are omitted in its equivalent direct graph as shown in Fig. 6.

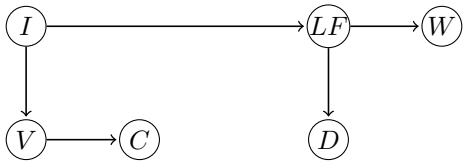

Fig. 6: An equivalent direct graph for the HCTBN of the COPD problem as given in Fig. 4; I: infection, V: sputum volume, C: cough, LF: lung function, D: dyspnea, W: wheeze.

Definition 4 (Equivalent Direct Models) *Let $\mathcal{N}$ be an HCTBN with intensity matrices $\mathbf{Q}$ and graph G. An equivalent direct model $\mathcal{M}$ is defined as a triple $\mathcal{M} = (G', \mathbf{Q}', P'_0)$ where graph $G' = (\mathbf{X}, \mathbf{Y}, \mathbf{E}')$ is defined as in Definition 3 and $\mathbf{Q}'$ is a set of intensity matrices over the nodes in graph G' and P'_0 is the initial distribution with $P'_0(X = 1) = 1$ or $P'_0(X = n_X + 1) = 1$, for any $X \in \mathbf{X}$.*

In addition, intensity matrices for any variable $Y \in \mathbf{Y}$ satisfy the following conditions:

- *If for any $X \in \mathbf{X}, X \notin \pi(Y)$, $Q^{\mathcal{M}}_{Y|\pi(Y)} = Q^{\mathcal{N}}_{Y|\pi(Y)}$ where $Q^{\mathcal{M}}_{Y|\pi(Y)}$ and $Q^{\mathcal{N}}_{Y|\pi(Y)}$ are the intensity matrices for variable Y in $\mathcal{M}$ and $\mathcal{N}$ respectively; otherwise, $Q^{\mathcal{M}}_{Y|\mathbf{K}=\mathbf{k}^{\mathcal{M}},\mathbf{K}'=\mathbf{k}'} = Q^{\mathcal{N}}_{Y|\mathbf{K}=\mathbf{k}^{\mathcal{N}},\mathbf{K}'=\mathbf{k}'}$, where $\mathbf{K} = \pi(Y) \cap \mathbf{X}$ and $\mathbf{K}' = \pi(Y) \cap \mathbf{Y}$, $\mathbf{k}^{\mathcal{M}}$ and $\mathbf{k}^{\mathcal{N}}$ are the values of*

variables $\mathbf{K}$ *in* $\mathcal{M}$ *and* $\mathcal{N}$ *respectively, and for any* $K \in \mathbf{K}$*, if* $k^{\mathcal{M}} \in \{1, \ldots n_K\}$*, where* n_K *is the number of possible values for variable* K *in* $\mathcal{M}$*, then* $k^{\mathcal{N}} = 1$*; otherwise* $k^{\mathcal{N}} = 2$*.*

Furthermore, for each variable $X \in \mathbf{X}$*, the intensity matrices* $Q^{\mathcal{M}}_{X|\pi(X)=\mathbf{u}}$ *are defined by re-ordering the states of* $Q^{\mathcal{N}}_{XH|\pi(X)\setminus\{H\}=\mathbf{u}}$ *from current indices* $1, \ldots, 2n_X$ *to* $1, 3, \ldots, 2n_X - 1, 2n_X, \ldots, 4, 2$*, where* $H = b(X)$ *in* $\mathcal{N}$*.*

Definition 4 gives a general procedure to transform the intensity matrices for variables $\mathbf{Y}$ and $\mathbf{X}$ from an HCTBN $\mathcal{N}$ to its equivalent direct model $\mathcal{M}$. The transformation involves two parts, one part for exponential variables $\mathbf{Y}$ and one part for hypoexponential variables $\mathbf{X}$. For variable $Y \in \mathbf{Y}$, its intensity matrices in $\mathcal{M}$ are simply a copy of those in $\mathcal{N}$ if its parents do not contain any hypoexponential variables. Otherwise, we also need to transform the values of hypoexponential variables from $\mathcal{M}$ to $\mathcal{N}$ before copying intensity matrices from $\mathcal{N}$. For example, variable X is an n_X-order hypoexponential variable and is a parent of exponential variable Y. Then the intensity matrices $Q^{\mathcal{M}}_{Y|X=1:n_X}$ for variable Y in $\mathcal{M}$ corresponds to $Q^{\mathcal{N}}_{Y|X=1}$ in $\mathcal{N}$, and $Q^{\mathcal{M}}_{Y|X=n_X+1:2n_X}$ to $Q^{\mathcal{N}}_{Y|X=2}$. For a hypoexponential variable X, its intensity matrices in $\mathcal{M}$ are transformed from $\mathcal{N}$ by amalgamating the joint intensity matrix of variable X and its associated hidden variable H and then reordering the resulting joint intensity matrix in a particular order.

For a binary variable X with n_X-order hypoexponential distribution in an HCTBN, the hypoexponential distribution is encoded in the associated n_X-valued hidden variable $H, H = b(X)$. In its equivalent direct model, the hypoexponential distribution is represented by adding additional states to its corresponding variable X'. Thus, the size of the state-space of variable X' grows to $2 \cdot n_X$ in the equivalent direct model from 2 states in the HCTBN.

Example 3 *For the HCTBN* $\mathcal{N}$ *as parameterized in Example 2, the structure of its equivalent Markov model* $\mathcal{M}$ *is given in Fig. 6 and we have an* 8×8 *intensity matrix for variable* LF *in* $\mathcal{M}$ *as given in the following:*

$$Q^{\mathcal{M}}_{LF} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -6 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -7 & 7 \\ 8 & 0 & 0 & 0 & 0 & 0 & 0 & -8 \end{pmatrix} \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array}$$

The size of the intensity matrix of variable LF *expands from* 2 *in* $\mathcal{N}$ *to* 8 *in* $\mathcal{M}$ *. In model* $\mathcal{M}$*, variable* LF *has eight states and states 1 to 4 correspond to state 1 for its associated variable in* $\mathcal{N}$*, and the rest states, i.e., 5 to 8, correspond to state 2.*

In addition, we also need to transform the intensity matrices for variable D*. Since variable* D *has the hypoexponential variable* LF *as a parent, we need to copy intensity matrices from* $\mathcal{N}$ *corresponding to each state of* LF *in* $\mathcal{M}$*, this gives:*

$$Q^{\mathcal{M}}_{D|LF\in\{1,2,3,4\}} = \begin{pmatrix} -29 & 29 \\ 30 & -30 \end{pmatrix} \quad Q^{\mathcal{M}}_{D|LF\in\{5,6,7,8\}} = \begin{pmatrix} -31 & 31 \\ 32 & -32 \end{pmatrix}$$

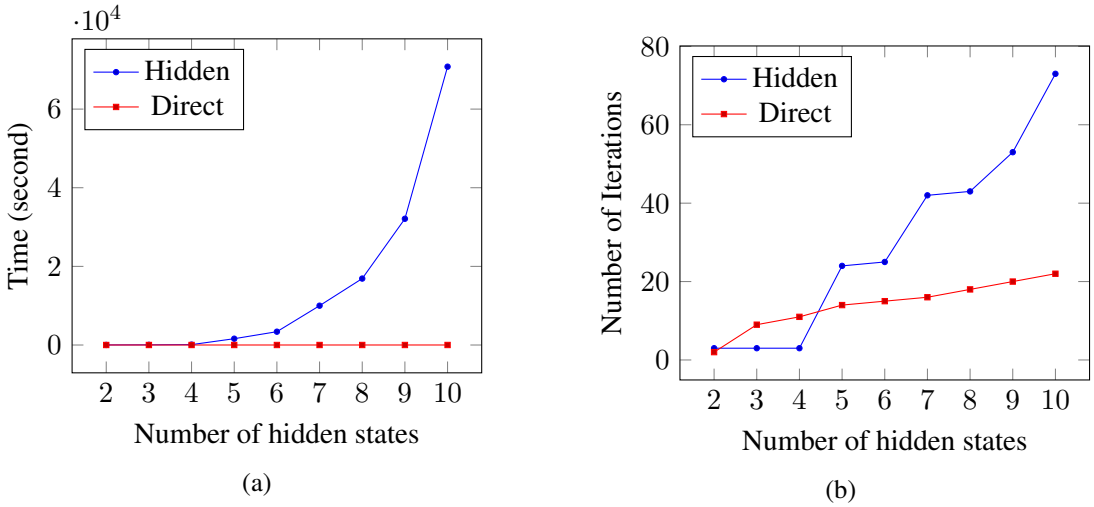

Fig. 7: Learning for COPD networks using the direct and hidden approaches with the number of hidden states ranging from 2 to 10 in the learned models, while it is fixed to 4 in the underlying model: the time until convergence in (a) and the number of iterations until convergence in (b).

7. Experiments

In the experiments, we studied parameter estimation in HCTBNs using the direct and hidden approaches. In both approaches, the EM algorithm is used to estimate parameters as either the hidden variables in HCTBNs are unobserved or the hidden states are unknown in their corresponding direct models. For the EM algorithm used in the methods, we compared the time and the number of iterations until the EM algorithm converges, and the quality of learned models in terms of log-likelihood. The model we used in the following experiments is given in Fig. 4, where we have two hypoexponential variables. For the hidden variables H_1 and H_2, the number of hidden states in the underlying model was both set to 4, and it varied from 2 to 10 in the learned models.

A number of software packages were used to learn parameters for HCTBNs. We used the existing CTBNs learning algorithms in the package *CTBN-RLE*[1] to estimate parameters in HCTBNs directly. For the direct approach, the transformation between a given HCTBN and its equivalent direct representation was implemented with R. We also employed *EMpht*[2] to learn parameters for this direct representation from *right censored data*, i.e., a variable staying in a state for at least a given amount of time. A more detailed discussion about censored data can be found in Gopalratnam et al. (2005). The synthetic training datasets for the COPD network consisted of approximately 6000 observations on average.

The time until the EM algorithm converges and the number of iterations are shown in Fig. 7a and Fig. 7b, respectively. The results in Fig. 7a suggest that the time until the EM algorithm converges grows exponentially with the number of states of the hidden variables using existing CTBNs learning algorithms, whereas there is little impact of the choice of the hidden states on the equivalent direct models (see the exponentially increasing time for HCTBNs when the number of hidden states increases from 6). Similarly, the number of iterations until convergence grows faster using existing CTBN learning algorithms than the equivalent direct models, whereas the difference is relatively

1. http://rlair.cs.ucr.edu/ctbnrle/
2. http://home.math.au.dk/asmus/pspapers.html

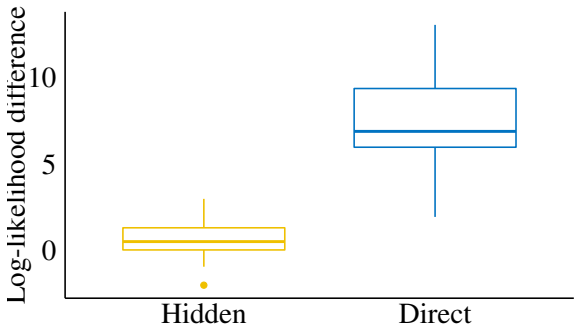

Fig. 8: Test log-likelihood for learned models using the direct and hidden approaches from 30 different underlying models. For a given underlying model, models are learned using each approach with 30 different starting parameters for the EM algorithms. The number of hidden states in both learned models and underlying models is set to 4. The log-likelihood difference is computed by subtracting the mean log-likelihood of learned models from those of the underlying models. Lower values indicate a better parameter estimation. Whether the difference was significant was computed by carrying out a paired t-test on the mean log-likelihood of the learned models using direct and hidden approaches with p-value less then 0.05.

small. By combining the results in Fig. 7a and Fig. 7b together, we can conclude that the time for each iteration in CTBNs grows exponentially with the number of hidden states. This is mainly attributed to the relatively large number of variables in the models, leading to a significant increase in the computation at each iteration.

We further evaluated the quality of the learned models in terms of log-likelihood. To rule out the randomness of starting parameters in the EM algorithms, we learned models with 30 different starting parameters for each approach. To test the performance for general models, these learning methods were used to estimate parameters for 30 underlying models with the same structure but different parameters. The test data were generated independently from learning training data and the log-likelihood difference was subtracted the mean log-likelihood on the test data of learned models from those of the underlying models, as shown in Fig. 8. A lower log-likelihood difference indicates higher quality of the learned models. It is clear that the hidden approach has a better and more stable estimation of parameters for the underlying models ($p < 0.05$ based on a paired $t-$test for the mean log-likelihood for both methods), which supports the conclusion that the quality of the learned models achieved by the hidden approach is significantly higher than that of the direct approach.

8. Conclusions

In this paper, we generalize our recently proposed HCTBN where a hypoexponential time distribution is restricted to a single hypoexponential variable by allowing any number of hypoexponential variables. In addition, we study the hidden and direct methods to estimate parameters from complete data. The experimental results show that the hidden approach indeed can learn models with significant higher quality. Transforming the learning task into equivalent direct models has the advantage of lower computational cost at the expense of the quality of learned models. Nevertheless, such an

approach will be infeasible when some non-hidden variables, i.e., exponential and hypoexponential variables, are partially known.

A limitation of HCTBNs so far is that the hypoexponential variables are restricted to be binary as the focus of this paper has been on introducing a richer time distribution. In future work, we aim to support multinomial hypoexponential variables, for example by using some states of the hidden variables for the state transitions of hypoexponential variables, and the rest for representing the hypoexponential time distribution. Using existing learning algorithms in CTBNs has the advantage of learning from partial trajectories, where observable variables are not fully known. However, these algorithms so far suffer from costly computation time. One possible solution to solve this problem is to decompose the EM algorithm into two parts, one part consisting of partially observed variables another part consisting of fully observed variables. This decomposition will significantly reduce the computation time, in particular when the exponential variables and their parents are fully observed.

References

N. T. J. Bailey. A statistical method of estimating the periods of incubation and infection of an infectious disease. *Nature*, 174(4420):139, 1954.

M. Bladt. A review on phase-type distributions and their use in risk theory. *ASTIN Bulletin*, 35(1): 145–161, 2005.

K. Gopalratnam, H. Kautz, and D. S. Weld. Extending continuous time Bayesian networks. In *Proceedings of the 20th National Conference on Artificial Intelligence*, volume 2, pages 981–986, 2005.

K. J. Gough. The estimation of latent and infectious periods. *Biometrika*, 64(3):559–565, 1977.

M. Liu, F. Stella, A. Hommersom, and P. J. F. Lucas. Representing hypoexponential distributions in continuous time Bayesian networks. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2018.

S. Moral, R. Rumí, and A. Salmerón. Mixtures of truncated exponentials in hybrid Bayesian networks. In *Proceedings of the 6th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, ECSQARU '01, pages 156–167, 2001.

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

U. Nodelman, C. R. Shelton, and D. Koller. Continuous time Bayesian networks. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, pages 378–387, 2002.

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

T. A. Seemungal, G. C. Donaldson, A. Bhowmik, D. J. Jeffries, and J. A. Wedzicha. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 161(5):1608–1613, 2000.

R. Verbelen. Phase-type distributions & mixtures of Erlangs. Master's thesis, 2013.