Continuous-Time Bayesian Networks with Clocks

Nicolai Engelmann [1] **Dominik Linzner** [1] **Heinz Koeppl** [1 2]

Abstract

Structured stochastic processes evolving in continuous time present a widely adopted framework to model phenomena occurring in nature and engineering. However, such models are often chosen to satisfy the Markov property to maintain tractability. One of the more popular of such memoryless models are Continuous Time Bayesian Networks (CTBNs). In this work, we lift its restriction to exponential survival times to arbitrary distributions. Current extensions achieve this via auxiliary states, which hinder tractability. To avoid that, we introduce a set of node-wise clocks to construct a collection of graph-coupled semi-Markov chains. We provide algorithms for parameter and structure inference, which make use of local dependencies and conduct experiments on synthetic data and a data-set generated through a benchmark tool for gene regulatory networks. In doing so, we point out advantages compared to current CTBN extensions.

1. Introduction

Dynamics on networks are abundant. Examples span from small scales as in gene regulatory networks (Stella et al., 2014) to large scales as in population or opinion dynamics on social networks (De et al., 2016). All these processes depend on some form of memory. Genes interact via expression of molecules with maturation times that are not exponentially distributed (Klann & Koeppl, 2012), neuron- and social dynamics depend on past events (Pillow et al., 2008).

While models with non-Markovian behavior exist, e.g. Hawkes processes (Linderman et al., 2016), they either

follow a fixed parametrization that hinders their expressiveness, or are defined in discrete time as Dynamic Bayesian Networks (Dagum et al., 1992) and are thus heavily over parametrized if no natural time scale exists. An expressive, but Markovian model in continuous time are Continuous Time Bayesian Networks (CTBNs) (Nodelman et al., 1995). There have been several attempts in extending CTBNs to model non-Markovian behavior (Nodelman et al., 2005; Gopalratnam et al., 2005; Liu et al., 2018). However, in these extensions, a CTBN has to be inflated with arbitrary many auxiliary states. This hinders the scalability of such models.

In this manuscript, we consider a different extension based on CTBNs, where we augment each local node of the network with its own clock process. This clock keeps track of the history of the node and determines the survival time of the nodes state. The dynamics of the clocks can be instantiated by arbitrary survival time distributions.

We derive augmented CTBNs with clocks as a novel generative model along with likelihoods for a complete observed process in section 2. In section 3, we show how the exact parameter and structure inference can be performed. Finally, in sections 4 and 5 we show not only tractable parameter and structure learning for two example parametric survival time distributions, but as well demonstrate at hand of controlled synthetic experiments that CTBNs can fail as structure classifiers if provided data is non-Markovian, while augmented CTBNs perform well.

1.1. Continuous Time semi-Markov and Markov Chains

We begin by laying out relevant key characteristics of homogeneous continuous time semi-Markov chains (CTSMC's) and explore their memoryless special case, continuous time Markov chains (CTMC's). A CTSMC is a non-Markovian jump processes with a renewal property evolving over continuous time, taking values in a countable state space $\mathcal{X}$. It can be completely described by a set of transition rates $\lambda : \mathcal{X} \times \mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0} \times \mathcal{X} \rightarrow \mathbb{R}_{\geqslant 0}$, where $\lambda(x, \tau, t; x')$ denotes the instantaneous rate of the process transitioning from state x to x' at time t if its last transition happened at time $t - \tau$. Using them, we define the exit rate $\lambda(x, \tau, t) \equiv \sum_{x' \neq x} \lambda(x, \tau, t; x')$. A CTSMC is homoge-

[1]Department of Electrical Engineering and Information Technology, Technische Universitat Darmstadt, Darmstadt, Germany [2]Department of Biology, Technische Universitat Darmstadt, Darmstadt, Germany. Correspondence to: Nicolai Engelmann <nicolai.engelmann@bcs.tu-darmstadt.de>.

neous if $\lambda(x, \tau, t; x') = \lambda(x, \tau; x')$ is dependent of t only via τ. In this work, we solely consider homogeneous CTSMC's in this sense. To generate a realization of the chain $\{X(t)\}_{t \geqslant 0}$, we have to first determine the time s from arrival in state x to the transition to another state, i.e. the survival time, from the cumulative distribution (Limnios & Oprisan, 2001)

$$F(s \mid x) \equiv P(\text{"survival time in state } x\text{"} \leqslant s) \quad (1)$$
$$= 1 - \exp\left(-\int_0^s d\tau \, \lambda(x, \tau)\right),$$

with the survival function $\Lambda(s \mid x) \equiv 1 - F(s \mid x)$. We then draw the next state x' conditioned on the last state x and the elapsed survival time s from a categorical distribution $x' \mid x, s \sim \text{Cat}(\lambda(x, s; x')/\lambda(x, s))$, where $x' \neq x$, $x, x' \in \mathcal{X}$. The property that the next state x' is drawn independently of the time s can be called time-direction independence (Maes et al., 2009). In this case, we define transition probabilities $\theta(x', x) \equiv \lambda(x, s; x')/\lambda(x, s)$ from state x to x'. The quantities $\lambda(x, s)$ and $\theta(x) \equiv \{\theta(x', x) \mid x' \neq x\}$ then characterize the process. Since we need to know the time of the last transition to draw the survival time and next state, the process in general is non-Markovian. However, we can augment the chain $\{X(t)\}_{t \geqslant 0}$ by a process $\{T(t)\}_{t \geqslant 0}$, for which $T(t) = \tau$ always keeps track of the time since the last transition, resetting back to 0 at each state change. We call $\{T(t)\}_{t \geqslant 0}$ the clock of the CTSMC. The two-component process $\{X(t), T(t)\}_{t \geqslant 0}$ on the product state space $\mathcal{X} \times \mathbb{R}_{\geqslant 0}$ is then homogeneous and Markovian (Limnios & Oprisan, 2001).

An important special case of a CTSMC is called a CTMC iff $\lambda(x, \tau; x') = \lambda(x; x')$ is constant for all τ. This results in the process $\{X(t)\}_{t \geqslant 0}$ becoming Markovian and memoryless on its own, independent of $\{T(t)\}_{t \geqslant 0}$. As a result $F(s \mid x) = 1 - \exp(-\lambda(x)s)$ becomes fixed to be an exponential distribution. Note, that a CTMC always satisfies time-direction independence.

1.2. Continuous Time Bayesian Networks

A CTBN is a collection of N homogeneous CTMCs $\{X_n(t) \mid t \geqslant 0, n \in \{1, \ldots, N\}\}$ evolving in parallel over a product state space $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_N$ (Nodelman et al., 1995). As a consequence, we can characterize the state of the CTBN $\boldsymbol{x} \in \mathcal{X}$ to be a collection of the states of its constituting CTMCs $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$. The processes of a CTBN are conditional on each other, with a dependency structure encoded in a directed graph $\mathcal{G} = (V, E)$ (which may contain cycles). Specifically, we associate vertices $v_n \in V$ with processes X_n and edges $(v_m, v_n) \in E$ with X_m conditioning X_n. We define the parent set of

the n'th node by $\text{par}_n \equiv \{v_m \mid (v_m, v_n) \in E\}$. We can then collect the processes of parenting nodes as $U_n(t) \equiv \{X_m(t) \mid m \in \text{par}_n\}$ taking values in $\mathcal{U}_n \equiv \times_{m \in \text{par}_n} \mathcal{X}_m$. We denote realizations of this (vector-valued) process by $\boldsymbol{u}_n \in \mathcal{U}_n$.

CTBNs are defined using the asynchrony of transitions in continuous time, which is that the probability of more than one jump at a given time instance scales in $o(h)$. Under this knowledge we can generate realizations of the CTBN by using the local quantities, node-wise conditional rates $\lambda_n : \mathcal{X}_n \times \mathcal{U}_n \to \mathbb{R}_{\geqslant 0}$ and node-wise conditional transition probabilities $\theta_n : \mathcal{X}_n \times \mathcal{X}_n \times \mathcal{U}_n \to [0, 1]$. Further, we can identify the global transition rate as the sum of node-wise rates $\lambda(\boldsymbol{x}) = \sum_{n=1}^N \lambda_n(x_n, \boldsymbol{u}_n)$. Thus, the time until the global state change $\boldsymbol{x} \to \boldsymbol{x}'$ takes place can be again determined via (1) in the form of an exponential distribution for a CTMC. We can then determine the next state $\boldsymbol{x}'$ given $\boldsymbol{x}$ by subsequently drawing the changing process X_n and its next state x_n' from categorical distributions with mass functions

$$p(n \mid \boldsymbol{x}) = \frac{\lambda_n(x_n, \boldsymbol{u}_n)}{\sum_m \lambda_m(x_m, \boldsymbol{u}_m)} \quad (2)$$
$$p(x_n' \mid \boldsymbol{x}, n) = \frac{\lambda_n(x_n, \boldsymbol{u}_n; x_n')}{\lambda_n(x_n, \boldsymbol{u}_n)}.$$

2. Bayesian Networks of Generic CTSMC's

2.1. Model Definiton

We define the augmented CTBN on a semi-continuous state space $\mathcal{X}_1 \times \mathbb{R}_{\geqslant 0} \times \mathcal{X}_2 \times \mathbb{R}_{\geqslant 0} \times \cdots \times \mathcal{X}_N \times \mathbb{R}_{\geqslant 0}$, with N local state variables $X_n(t)$ and local clocks $T_n(t)$ for $n \in \{1, \ldots, N\}$. Like above, we express this collection in a vector $(\boldsymbol{x}, \boldsymbol{\tau}) = (x_1, x_2, \ldots, x_N, \tau_1, \tau_2, \ldots, \tau_N)$. We refer to the random variable associated with the collection of all local state variables by $\boldsymbol{X}(t)$ and the one associated with all local clocks by $\boldsymbol{T}(t)$. Like in original CTBN's, we can argue, that under knowledge of all (generalized) states, the global process obeys the Markov property. We fully inherit the semantics of a graph inducing the dependencies from CTBN's and allow a conditioning only via the discrete state components in $\boldsymbol{X}(t)$ like in other proposed CTBN extensions (Nodelman et al., 2005; Gopalratnam et al., 2005; Liu et al., 2018). Thus, a process being conditionally dependent on another process in the network is only dependent on its discrete state component and not its clock, cf. fig. 1 a).

We introduce local conditional transition rates $\lambda(x_n, \tau_n, \boldsymbol{u}_n; x_n')$ for a process $X_n(t)$ depending on its state $X_n(t) = x_n$, its clock $T_n(t) = \tau_n$ and the states of other processes $U_n(t) = \boldsymbol{u}_n$ referred to as parent state like above. These rates are completely determined by our choice of the survival time distributions. We show this
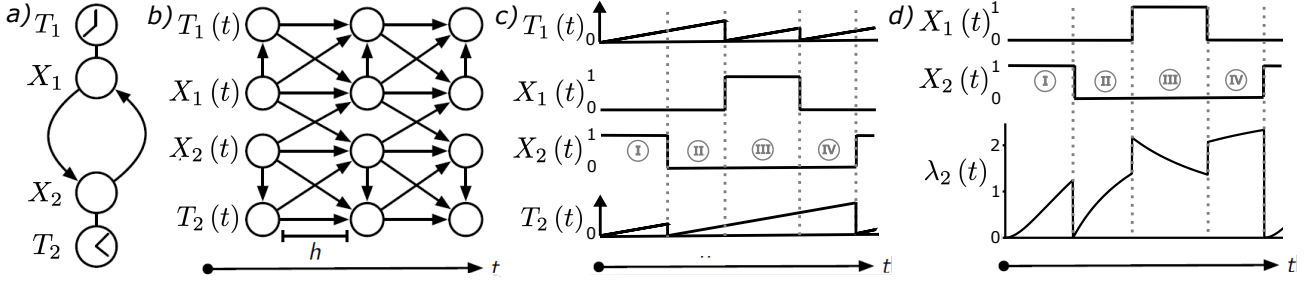
Figure 1. A graphic illustrating the semantic construction of the augmented CTBN. a) shows $N = 2$ nodes with their local clocks. b) depicts their conditional relationships in a small time window h as an equivalent Dynamic Bayesian Network. c) shows a sample trajectory for binary X_1 and X_2 highlighting the four classes of time-windows for X_2: full (I), censored (II), truncated and censored (III), and truncated (IV). d) gives an exemplary time-course for the transition rate of X_2 with Gamma and Weibull survival time distributions.

in Appendix A.3. Like above, we can define local exit rates $\lambda(x_n, \tau_n, \boldsymbol{u}_n) = \sum_{x'_n \neq x_n} \lambda(x_n, \tau_n, \boldsymbol{u}_n; x'_n)$. Similarly to CTBN's, if the associated graph $\mathcal{G}$ has an edge (v_m, v_n), then $X_m(t) \in U_n(t)$ is a parent of $X_n(t)$. We also show in Appendix A.3, that using the Markov property of the global process $\{\boldsymbol{X}(t), \boldsymbol{T}(t)\}$, the asynchronous transition condition and the inherited properties of the constituting CTSMC's are sufficient to fully characterize the global survival time distribution and transition probabilities in terms of the local conditional transition rates of the processes. We do so by integrating the probabilities of all constituting processes keeping their current states x_n over small time windows of duration h and thus find the survival function of the global process with respect to $\boldsymbol{X}(t)$ in the limit $h \to 0$ by

$$\Lambda(s \,|\, \boldsymbol{x}, \, \boldsymbol{\tau}) = \exp\left(-\sum_{n=1}^{N} \int_{\tau_n}^{\tau_n + s} \mathrm{d}\sigma \, \lambda(x_n, \, \sigma, \, \boldsymbol{u}_n)\right)$$

$$= \prod_{n=1}^{N} \frac{\Lambda_n(s + \tau_n \,|\, x_n, \, \boldsymbol{u}_n)}{\Lambda_n(\tau_n \,|\, x_n, \, \boldsymbol{u}_n)}, \qquad (3)$$

in terms of local survival functions $\Lambda_n(s \,|\, x_n, \, \boldsymbol{u}_n) = \exp\left(-\int_0^s \mathrm{d}\sigma \, \lambda(x_n, \, \sigma, \, \boldsymbol{u}_n)\right)$ conditioned on the parents' state $\boldsymbol{u}_n$. An illustration of this approach is given in fig. 1 b). Consequently, introducing the global $F(s \,|\, \boldsymbol{x}, \, \boldsymbol{\tau}) = 1 - \Lambda(s \,|\, \boldsymbol{x}, \, \boldsymbol{\tau})$ and local $F_n(s \,|\, x_n, \, \boldsymbol{u}_n) = 1 - \Lambda_n(s \,|\, x_n, \, \boldsymbol{u}_n)$ cumulative distribution functions along with their local density functions $f_n(s \,|\, x_n, \, \boldsymbol{u}_n)$, we can give the density function of the global survival time by

$$p(s \,|\, \boldsymbol{x}, \boldsymbol{\tau}) = \frac{\mathrm{d}F(s \,|\, \boldsymbol{x}, \, \boldsymbol{\tau})}{\mathrm{d}s} \qquad (4)$$

$$= \sum_{n=1}^{N} \frac{f_n(s + \tau_n \,|\, x_n, \, \boldsymbol{u}_n)}{\Lambda_n(\tau_n \,|\, x_n, \, \boldsymbol{u}_n)} \prod_{k \neq n} \frac{\Lambda_k(s + \tau_k \,|\, x_k, \, \boldsymbol{u}_k)}{\Lambda_k(\tau_k \,|\, x_k, \, \boldsymbol{u}_k)}.$$

Note, that we can describe $p(s \,|\, \boldsymbol{x}, \boldsymbol{\tau})$ in terms of only a finite set of local distributions, which are themselves independent of the current clock values $\boldsymbol{\tau}$. The query for the next global state $\boldsymbol{x}'$ can be answered by subsequently drawing the changing process X_n and its next state x'_n given the clock values $\boldsymbol{\tau}$ and the duration s from categorical distributions with mass functions

$$p(n \,|\, \boldsymbol{x}, \boldsymbol{\tau}, s) = \frac{\lambda_n(x_n, \, \tau_n + s, \, \boldsymbol{u}_n)}{\sum_m \lambda_m(x_m, \, \tau_m + s, \, \boldsymbol{u}_m)} \qquad (5)$$

$$p(x'_n \,|\, \boldsymbol{x}, \boldsymbol{\tau}, s, n) = \frac{\lambda_n(x_n, \, \tau_n + s, \, \boldsymbol{u}_n; \, x'_n)}{\lambda_n(x_n, \, \tau_n + s, \, \boldsymbol{u}_n)}.$$

As a result, we update the clock values $\tau_n \to 0$ for the changing process and $\forall m \neq n : \tau_m \to \tau_m + s$ for the others. We provide a detailed derivation of the global survival time density and the transition probabilities in Appendix A.3. We are now in the position to generate CTBNs with any parametrized distribution for any state x_n and condition $\boldsymbol{u}_n$ for all n. It is no coincidence, that (4) resembles the density of a minimum distribution, which renders trajectory sampling straight forward and efficient for parametric local survival time distributions despite having to evaluate truncated distributions. This can be done with a Gillespie algorithm (outlined in Appendix C), subsequently drawing global survival times, changing processes and next states. After each step, we have to update $\boldsymbol{x}$ and $\boldsymbol{\tau}$ accordingly. Note, that we share the occurrence of truncations with generalized semi-Markov chains, in which triggered events also "interrupt" the survival procedure and we have to "continue" bearing truncated distributions (Cassandras & Lafortune, 2006).

2.1.1. PARAMETRIZATION OF LOCAL SURVIVAL TIME DISTRIBUTIONS

In (1) from section 1.1 we can see, that the exit rates $\lambda(x, s)$ can be determined if $\Lambda(s \,|\, x)$ is known. This also

holds for the local $\lambda_n\left(x_n,\, s,\, \boldsymbol{u}_n\right)$ given $\Lambda_n\left(s\,|\,x_n,\, \boldsymbol{u}_n\right)$ in the augmented CTBN. Despite being able to parametrize local distributions by transition rate functions directly, many distributions are conveniently parametrized by a different set of quantities. In consideration of possible applications, we therefore introduce hyperparameters $\boldsymbol{\phi} \equiv \left\{\forall n, x_n, \boldsymbol{u}_n : \phi_n\left(x_n,\, \boldsymbol{u}_n\right)\right\}$ representing a chosen distribution per process n, its state x_n and parent state $\boldsymbol{u}_n$, and consequently let $\lambda_n\left(x_n,\, s,\, \boldsymbol{u}_n\right)$ and $\Lambda_n\left(s\,|\,x_n,\, \boldsymbol{u}_n\right)$ take on a functional form dependent on $\phi_n\left(x_n,\, \boldsymbol{u}_n\right)$. We also restrict our considerations to cases, where the local processes in the augmented CTBN obey time-direction independence between state changes, which allows the factorization $\lambda_n\left(x_n,\, \tau_n,\, \boldsymbol{u}_n;\, x'_n\right) \equiv \theta_n\left(x'_n\,|\,x_n,\, \boldsymbol{u}_n\right) \lambda_n\left(x_n,\, \tau_n,\, \boldsymbol{u}_n\right)$. As a consequence, $\boldsymbol{\phi}$ suffices to describe all survival time distributions. To account for this, we write $\lambda_n\left(x_n,\, s,\, \boldsymbol{u}_n\right) \equiv \lambda_n\left(s,\, \phi\left(x_n,\, \boldsymbol{u}_n\right)\right)$ and $\Lambda_n\left(s\,|\,x_n,\, \boldsymbol{u}_n\right) \equiv \Lambda_n\left(s\,|\,\phi\left(x_n,\, \boldsymbol{u}_n\right)\right)$. To give a short example, for a Weibull distribution in shape-rate parametrization $(k,\, b)$, we would have the free parameters $\phi_n\left(x_n,\, \boldsymbol{u}_n\right) \equiv \left(k_n\left(x_n,\, \boldsymbol{u}_n\right), b_n\left(x_n,\, \boldsymbol{u}_n\right)\right)$ for local state x_n and parent state $\boldsymbol{u}_n$, giving

$$\Lambda_n\left(s\,|\,\phi_n\left(x_n,\, \boldsymbol{u}_n\right)\right) \tag{6}$$
$$= \exp\left(-b_n\left(x_n,\, \boldsymbol{u}_n\right) s^{k_n\left(x_n,\, \boldsymbol{u}_n\right)}\right)$$
$$\lambda_n\left(s,\, \phi_n\left(x_n,\, \boldsymbol{u}_n\right)\right)$$
$$= b_n\left(x_n,\, \boldsymbol{u}_n\right) k_n\left(x_n,\, \boldsymbol{u}_n\right) s^{k_n\left(x_n,\, \boldsymbol{u}_n\right)-1}.$$

2.2. Path Measure

The global path measure of our augmented CTBN can be derived from the expressions in (4) and (5), and factorizes in terms of its discrete $\boldsymbol{X}\left(t\right)$ and continuous $\boldsymbol{T}\left(t\right)$ components for single transitions as a consequence of our simplifications in section 2.1.1. For a fixed number M of transitions it is given by a product of M global density function evaluations (4) and M evaluations of the mass functions representing the categorical transition directions (5). Shown in detail in Appendix A.4, we obtain the density for a single time-window from entry in state $\boldsymbol{x}$ with clock configuration $\boldsymbol{\tau}$ on to our transition to $\boldsymbol{x}'$ after duration s with the n-th process changing by

$$p\left(\boldsymbol{x}',\, \boldsymbol{\tau}'\,|\,\boldsymbol{x},\, \boldsymbol{\tau},\, \boldsymbol{\theta},\, \boldsymbol{\phi}\right) = p\left(\boldsymbol{x}'\,|\,\boldsymbol{x},\, \boldsymbol{\theta}\right) p\left(\boldsymbol{\tau}'\,|\,\boldsymbol{x}',\, \boldsymbol{x},\, \boldsymbol{\tau},\, \boldsymbol{\phi}\right)$$
$$= \theta_n\left(x'_n\,|\,x_n,\, \boldsymbol{u}_n\right) \frac{f_n\left(s + \tau_n\,|\,\phi_n\left(x_n,\, \boldsymbol{u}_n\right)\right)}{\Lambda_n\left(\tau_n\,|\,\phi_n\left(x_n,\, \boldsymbol{u}_n\right)\right)}$$
$$\times \prod_{k \neq n} \frac{\Lambda_k\left(s + \tau_k\,|\,\phi_k\left(x_k,\, \boldsymbol{u}_k\right)\right)}{\Lambda_k\left(\tau_k\,|\,\phi_k\left(x_k,\, \boldsymbol{u}_k\right)\right)}, \tag{7}$$

while the factor $p\left(\boldsymbol{x}'\,|\,\boldsymbol{x},\, \boldsymbol{\theta}\right)$ is independent of the clocks and the survival times. We recognize this from embedded Markov chains in the continuous time context (Pinsky & Karlin, 2010). Further, we must specify

$$\boldsymbol{x}' = \left(x_1,\, \ldots,\, x_{n-1},\, x'_n,\, x_{n+1},\, \ldots,\, x_N\right)$$
$$\boldsymbol{\tau}' = \left(\tau_1 + s,\, \ldots,\, \tau_{n-1} + s,\, 0,\, \tau_{n+1} + s,\, \ldots,\, \tau_N + s\right).$$

Since successive time-windows condition on $\left(\boldsymbol{x}',\, \boldsymbol{\tau}'\right)$ as the new "starting point", we can construct the measure for M transitions by a product of M such time-windows. It is also reasonable to assume the final window to be fully censored, so that there is no transition exactly at the end. In this case, the measure for this time-window reduces to an evaluation of the global survival function (3). Consequently, gathering $\mathbf{X} = \left\{\boldsymbol{x}^{(0)},\, \boldsymbol{x}^{(1)},\, \ldots,\, \boldsymbol{x}^{(M)}\right\}$ and $\mathbf{T} = \left\{\boldsymbol{\tau}^{(0)},\, \boldsymbol{\tau}^{(1)},\, \ldots,\, \boldsymbol{\tau}^{(M)}\right\}$, we can write

$$p\left(\mathbf{X},\, \mathbf{T}\,|\,\boldsymbol{\theta},\, \boldsymbol{\phi}\right) = p\left(\mathbf{X}\,|\,\boldsymbol{\theta}\right) p\left(\mathbf{T}\,|\,\mathbf{X},\, \boldsymbol{\phi}\right) \tag{8}$$
$$= \prod_{m=0}^{M-1} p\left(\boldsymbol{x}^{(m+1)}\,|\,\boldsymbol{x}^{(m)},\, \boldsymbol{\theta}\right)$$
$$\times \prod_{m=0}^{M-1} p\left(\boldsymbol{\tau}^{(m+1)}\,|\,\boldsymbol{x}^{(m+1)},\, \boldsymbol{x}^{(m)},\, \boldsymbol{\tau}^{(m)},\, \boldsymbol{\phi}\right).$$

which can be made explicit by directly inserting the respective terms outlined in (7). We give the factors of the path measure associated with the survival times for Weibull and Gamma distributions in section 4 and Appendix A.5. Under knowledge of (8), we can work on the question of inference of model parameters of the augmented CTBN. Before we do so, however, we want to briefly address related work.

2.3. Related Work

In the following we want to argue about the design choices of our model and compare them to models existing in literature. This is a discussion about different clock augmentations. Proposed extensions to CTBNs using phase-type distributions (Gopalratnam et al., 2005)(Nodelman et al., 2005) and hidden nodes (Liu et al., 2018) allow per-process survival time related structures. In both models, multiple so-called phases keep track of the survival time and are associated to a so-called generalized state and only this generalized state is relevant for the conditions imposed on a dependent process. The CTBN extension based on hidden nodes in addition only allows conditioning of nodes on binary state spaces. This limits the models applicability and requires comparably large networks to model nodes with many states and arbitrary non-exponential survival time distributions.

While phase-type CTBNs are similar in terms of expressiveness, they suffer from two distinct problems which we aim to avoid. The first problem is limited tractability due to the introduction of auxiliary states, the second being the heavy over-parametrization of phase-type distributions (Hobolth et al., 2019). We alleviate both problems by replacing arbitrary auxiliary states with, also arbitrary, survival time distributions with few parameters.

3. Inference of Model Parameters

Model inference can be done with varying efficiency, depending on the choice of the survival time distribution. We can see in (8), that the likelihood factors not only in single time-window terms, but also in transition probabilities $\boldsymbol{\theta}$ and survival time distribution parameters $\boldsymbol{\phi}$. We can therefore give the posterior for independent $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ by

$$
\begin{aligned}
p\left(\boldsymbol{\theta}, \boldsymbol{\phi} \,|\, \mathbf{X}, \mathbf{T}\right) &\propto p\left(\mathbf{X}, \mathbf{T} \,|\, \boldsymbol{\theta}, \boldsymbol{\phi}\right) p\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{\phi}\right) \\
&= p\left(\mathbf{X} \,|\, \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right) p\left(\mathbf{T} \,|\, \mathbf{X}, \boldsymbol{\phi}\right) p\left(\boldsymbol{\phi}\right) \\
&\propto p\left(\boldsymbol{\theta} \,|\, \mathbf{X}\right) p\left(\boldsymbol{\phi} \,|\, \mathbf{X}, \mathbf{T}\right).
\end{aligned} \tag{9}
$$

As a consequence, we can infer $\boldsymbol{\theta}$ traditionally in terms of transition probabilities of an embedded Markov chain (Pinsky & Karlin, 2010). This is well studied, and can be done analytically using a conjugate Dirichlet prior, so we refer to the original literature (Nodelman et al., 2003) for the expressions. In (7), we can also see, that each single time-window only depends on parameters $\theta_n\left(x_n' \,|\, x_n, \boldsymbol{u}_n\right)$ and $\phi_n\left(x_n, \boldsymbol{u}_n\right)$ associated with the relevant local x_n and parent state $\boldsymbol{u}_n$ configurations. If we assume all $\phi_n\left(x_n, \boldsymbol{u}_n\right)$ to be mutually independent, $p\left(\boldsymbol{\phi} \,|\, \mathbf{X}, \mathbf{T}\right)$ also factors in the local marginal posteriors $p\left(\phi_n\left(x_n, \boldsymbol{u}_n\right) \,|\, \mathbf{X}, \mathbf{T}\right)$, significantly reducing the curse of dimensionality in associated inference and learning problems. In this case, the marginal posterior $p\left(\phi_n\left(x_n, \boldsymbol{u}_n\right) \,|\, \mathbf{X}, \mathbf{T}\right)$ can be determined via

$$
\begin{aligned}
&p\left(\phi_n\left(x_n, \boldsymbol{u}_n\right) \,|\, \mathbf{X}, \mathbf{T}\right) \\
&\propto \prod_{m=1}^{M} \frac{f_n\left(s^{(m)} + \tau_n^{(m)} \,\Big|\, \phi_n\left(x_n, \boldsymbol{u}_n\right)\right)^{1-c(m, n, x_n, \boldsymbol{u}_n)}}{\Lambda_n\left(\tau_n^{(m)} \,\Big|\, \phi_n\left(x_n, \boldsymbol{u}_n\right)\right)} \\
&\quad \times \Lambda_n\left(s^{(m)} + \tau_n^{(m)} \,\Big|\, \phi_n\left(x_n, \boldsymbol{u}_n\right)\right)^{c(m, n, x_n, \boldsymbol{u}_n)} \\
&\quad \times p\left(\phi_n\left(x_n, \boldsymbol{u}_n\right)\right),
\end{aligned} \tag{10}
$$

where $s^{(m)} = \max\left\{\boldsymbol{\tau}^{(m+1)} - \boldsymbol{\tau}^{(m)}\right\}$ is the duration of the m-th time-window and $c\left(m, n, x_n, \boldsymbol{u}_n\right) = 1$ if it did not end with a transition of X_n, i.e. it is censored, otherwise 0. Note, that if $\tau_n^{(m)} = 0$, the denominator is

1, hence the time-window is not truncated for X_n. An illustration of the four possible cases of censoring and truncation can be seen in fig. 1 c). If we omit the prior on the rhs of (10), we obtain the factor of the likelihood for $\phi_n\left(x_n, \boldsymbol{u}_n\right)$. By inserting our example (6) from section 2.1.1 and recognizing $f_n\left(s \,|\, \phi_n\left(x_n, \boldsymbol{u}_n\right)\right) = \lambda_n\left(s, \phi_n\left(x_n, \boldsymbol{u}_n\right)\right) \Lambda_n\left(s \,|\, \phi_n\left(x_n, \boldsymbol{u}_n\right)\right)$, we immediately obtain the expression for the Weibull case.

If we assume not a fixed graph $\mathcal{G}$ imposing structural dependencies in the augmented CTBN but a distribution over graphs $p\left(\mathcal{G}\right)$, we render the parent sets par_n (and the constitution of U_n) and consequently the $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ dependent of $\mathcal{G}$, and can find the posterior of $\mathcal{G}$ by

$$
\begin{aligned}
p\left(\mathcal{G} \,|\, \mathbf{X}, \mathbf{T}\right) &\propto p\left(\mathbf{X} \,|\, \mathcal{G}\right) p\left(\mathbf{T} \,|\, \mathbf{X}, \mathcal{G}\right) p\left(\mathcal{G}\right) \tag{11} \\
&= \left(\int_{\boldsymbol{\Theta}} \mathrm{d}\boldsymbol{\theta}\, p\left(\mathbf{X} \,|\, \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta} \,|\, \mathcal{G}\right) \right. \\
&\quad \left. \times \int_{\boldsymbol{\Phi}} \mathrm{d}\boldsymbol{\phi}\, p\left(\mathbf{T} \,|\, \mathbf{X}, \boldsymbol{\phi}\right) p\left(\boldsymbol{\phi} \,|\, \mathcal{G}\right) \right) p\left(\mathcal{G}\right),
\end{aligned}
$$

where again, while the leftmost integral is analytically solvable, the rightmost can be reduced to a product of low-dimensional integrations under the mentioned assumption that the $\phi_n\left(x_n, \boldsymbol{u}_n\right)$ are independent. Then, the term becomes a product of integrals over the single $\phi_n\left(x_n, \boldsymbol{u}_n\right)$ with the integrals containing the likelihood term from the rhs of (10). In addition to that, our approach preserves structure modularity. Consequently, as long as the prior $p\left(\mathcal{G}\right) = \prod_{n=1}^{N} p\left(\mathcal{G}_n\right)$ factors in the same way, where $\mathcal{G}_n = (V, E_n)$ consists of only the incoming edges to process X_n, we can calculate the posterior of the full graph via $p\left(\mathbf{X}, \mathbf{T} \,|\, \mathcal{G}\right) p\left(\mathcal{G}\right) = \prod_{n=1}^{N} p\left(\mathbf{X}, \mathbf{T} \,|\, \mathcal{G}_n\right) p\left(\mathcal{G}_n\right)$ and then $\mathcal{G} = (V, \bigcup_n E_n)$, even more reducing dimensionality.

Because the augmented CTBN preserves the properties layed out in (9) and (11), it is - in contrast to existing CTBN extensions - possible to perform exact inference fully analytical for some non-exponential distributions. In Appendix A.6 we give the analytical marginal likelihoods (11) and posteriors (9) along with their conjugate priors and sufficient statistics for an augmented CTBN with Rayleigh survival time distributions as an example.

4. Example Survival Time Distributions

The description of the augmented CTBN given above is valid for a broad family of parametric and non-parametric survival time distributions. Before we conduct experiments, we want to give two specific implementations of the augmented CTBN based on Weibull and Gamma distributions. Those are not only important in survival analysis (D.R. Cox, 1984) and queueing theory, but also exhibit distinct numeri-

cal properties. In addition to that, Weibull CTBNs have a non-exponential tail in the general case and are therefore naturally hard to approximate by PH distributions used in related publications. Gamma r.v.'s on the other hand can be built by sums of exponential r.v.'s in the case of integer shape parameters. To give an intuition, we illustrate an exemplary rate progression for a CTBN with Gamma and Weibull distributions in fig. 1 d).

4.1. Weibull CTBN's

For the Weibull distribution, we chose the parametrization outlined in section 2.1.1 (see (6). To keep our notation simple, we omit the $(x, \boldsymbol{u})$ in the following, if no ambiguity occurs. Consequently, the factors of the likelihood associated with only the survival time parameters $\phi_n(x_n, \boldsymbol{u}_n) \equiv (k_n, b_n)$ then read

$$p(k_n, b_n \mid \mathbf{X}, \mathbf{T})$$

$$\propto p(k_n, b_n) b_n^{|S_f|} k_n^{|S_f|} \left(\prod_{m=1}^{|S_f|} s_{f,m} \right)^{k_n - 1}$$

$$\exp\left(-b_n \left(\sum_{m=1}^{|S_f|} s_{f,m}^{k_n} + \sum_{m=1}^{|S_c|} s_{c,m}^{k_n} - \sum_{m=1}^{|S_t|} s_{t,m}^{k_n} \right) \right),$$

where the set $S_f = \{s_{f,1}, s_{f,2}, \dots\}$ contains all values, the clock would have displayed just before the considered process transitioned, i.e. $\tau_n + s$. Furthermore, S_c gathers all clock samples, at which a parent change happens, while S_t gathers the samples at the beginning of the time-windows, i.e. the truncation times. We outline the construction with focus on sample efficiency in Appendix A.5. An advantage of the Weibull distribution is its exponential form and monomial rates. This allows us to evaluate sums of monomials in a logarithmic domain directly, which results in an improved numerical stability, especially compared with a Gamma distribution. Another interesting property of the distribution is its expressibility of sub- and superexponential tails.

4.2. Gamma CTBN's

The parameterization of the Gamma distributions in our CTBN is chosen by a rate α and shape β parameter similar to the Weibull case. The local survival functions of the CTBN for the n-th process can then be given by $\Lambda_n(s \mid \phi_n(x_n, \boldsymbol{u}_n)) = \frac{\Gamma(\alpha_n(x, \boldsymbol{u}), \beta_n(x, \boldsymbol{u})s)}{\Gamma(\alpha_n(x, \boldsymbol{u}))}$, where the Γ in the numerator denotes the upper incomplete gamma function. We can retrieve the exit rates via $\lambda_n(s, \phi_n(x_n, \boldsymbol{u}_n)) \Gamma(\alpha_n(x, \boldsymbol{u}), \beta_n(x, \boldsymbol{u})s) = \beta_n(x, \boldsymbol{u})^{\alpha_n(x, \boldsymbol{u})} s^{\alpha_n(x, \boldsymbol{u})-1} \exp(-\beta_n(x, \boldsymbol{u})s)$. The likelihood associated with the Gamma distribution can be

found in Appendix A.5. Numerically disadvantageous, the logarithm of the incomplete gamma function is comparably hard to obtain and if evaluated in the linear domain first, a wide range of parameter values for different $(x, \boldsymbol{u})$ can cause the likelihood to contain ratios with numerators and denominators close to or below machine precision.

5. Experimental Work

In this section, we give simulations for parameter and structure inference based on synthetic trajectories from augmented CTBNs generated using the Gillespie sampling procedure mentioned in section 2.1. In the end, we highlight a result for structure inference of an augmented CTBN under a Gamma distribution assumption on trajectories obtained through GeneNetWeaver (Schaffter et al., 2011) for a gene regulatory network.

5.1. Parameter and Structure Inference on Synthetic Trajectories

For this scenario, we sampled trajectories of varying lengths from random augmented CTBN's with four nodes via the Gillespie algorithm from Appendix C and performed parameter and structure inference on the trajectories. First, we describe, how the random graph including the random survival time distribution parameters is sampled. From a uniform degree distribution we determined the size $|\mathcal{U}_n|$ of the independent parent set for node n. Subsequently we drew a sample from a *uniform* Dirichlet distribution to obtain a random categorical distribution. To determine the nodes in the parent set, we then chose the $|\mathcal{U}_n|$ maximum values in the obtained vector. We did so for each node to obtain random independent parent sets. Then, for each independent parameter set in the CTBN associated with the sampled graph, we sampled random parameters for the Weibull and Gamma distributions again from Gamma distributions. Their shapes and rates (Gamma: $\alpha_1 = 40, \beta_1 = 5$ for the shape and $\alpha_2 = 25, \beta_2 = 2.5$ for the rate generation, Weibull: $\alpha_1 = 8, \beta_1 = 0.5$ for the shape and $\alpha_2 = 5, \beta_2 = 3$ for the rate) were chosen to lead to reasonable model parameters with a significant variance in the samples and pronounced non-exponential appearance.

5.1.1. PARAMETER INFERENCE

For parameter inference, we carry out an experiment for Weibull and Gamma CTBN's each. We give a plot of the mean squared error (MSE) showing quantiles over the $1\,000$ trajectories in fig. 2. For each of the CTBN's, we chose a random graph and sampled $1\,000$ trajectories with a size of $10\,000$ transitions each from it. Note, that the amount of transitions corresponds to the global process. We therefore obtained a significantly smaller amount of samples per parameter set $\phi_n(x_n, \boldsymbol{u}_n)$ of a single node, local state and
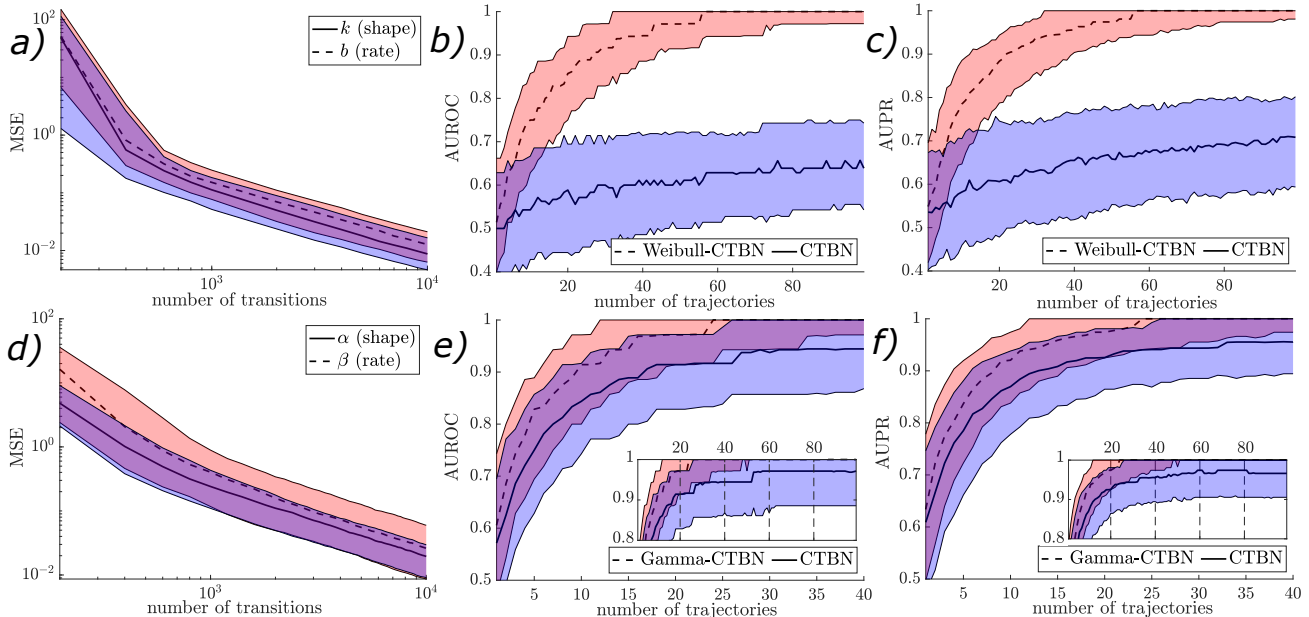
Figure 2. a) mean squared error (MSE) distribution for varying numbers of transitions in the Weibull parameter inference scenario. 80% of the calculated MSE's are bounded by the colored area. The blue area corresponds to the shape and the red area to the rate. An exponential convergence to the true parameter sets can be seen as a straight line in the dual log plot. b) & c) AUROC and AUPR for the Weibull structure inference scenario for varying numbers of trajectories compared with a traditional CTBN. With a mean shape of around $k \approx 15$ for the generated Weibull distributions, the traditional CTBN slowly approaches the true graph, but fails to classify the ground truth with the given number of trajectories. d) MSE similar to a) for the Gamma parameter inference scenario. e) & f) AUROC and AUPR for the Gamma structure inference scenario similar to the Weibull case. A generally faster convergence (mean shape around $\alpha \approx 8$) compared to Weibull can be seen and although the traditional CTBN fails to classify the graph, it approaches it closely (see inlet plot).

parent state. Then, we inferred the parameters for each graph given different numbers of transitions reaching to the maximum of 10 000. For the prior, we chose a uniform prior imposing box constraints from 0.1 to 100 for both parameters. Since we could only perform numerical posterior updates on a discrete grid, we chose to always evaluate a single likelihood to obtain an arbitrary close approximation of the posterior mode in this experiment. To calculate the MAP-estimates, we employed a limited-memory BFGS algorithm using projected gradients to stay in the bounds dictated by the uniform prior. We did so for each of the 1 000 trajectories and then calculated the MSE of each parameter estimate for each inferred number of time-windows over all trajectories. An exponential convergence to the true parameter values can be observed in the plots in fig. 2.

5.1.2. STRUCTURE INFERENCE

We performed the structure inference experiment for Gamma and Weibull CTBN's separately as well and compared the results to traditional CTBN's. For each, we generated 500 different random graphs. The following procedure was performed for each of those graphs. We sampled 100 trajectories corresponding to a time-window of 5 units

each (roughly 20 to 30 transitions). We chose a uniform prior over the graphs and again the uniform prior from the parameter inference above for those. For each of the trajectories containing the 100 time-windows, we calculated the marginal likelihood (11) of the graph via sample-based numerical integration employing a Romberg integration and performed a posterior update for the respective survival time ϕ and jump parameters θ. Using the calculated marginal likelihood, we did a posterior update for the graph. We saved the result and proceeded in the same way with the other 99 trajectories until we had 100 subsequent posterior updates for parameters and graph. The posterior updates for the parameters were dumped afterwards. This has been carried out with all the 500 graphs, so that we sampled 100 posterior distributions for each multiple of 5 units of time over the random graphs. For each of the graph posteriors, we gathered the marginal probabilities of the presence of an edge in a weighted adjacency matrix. Using the true graph and the matrices, we then calculated the AUROC and AUPR classification scores. For the structure inference for CTBNs we followed the original paper (Nodelman et al., 2003). We can see in fig. 2, that traditional CTBN's performed surprisingly well when confronted with samples from the
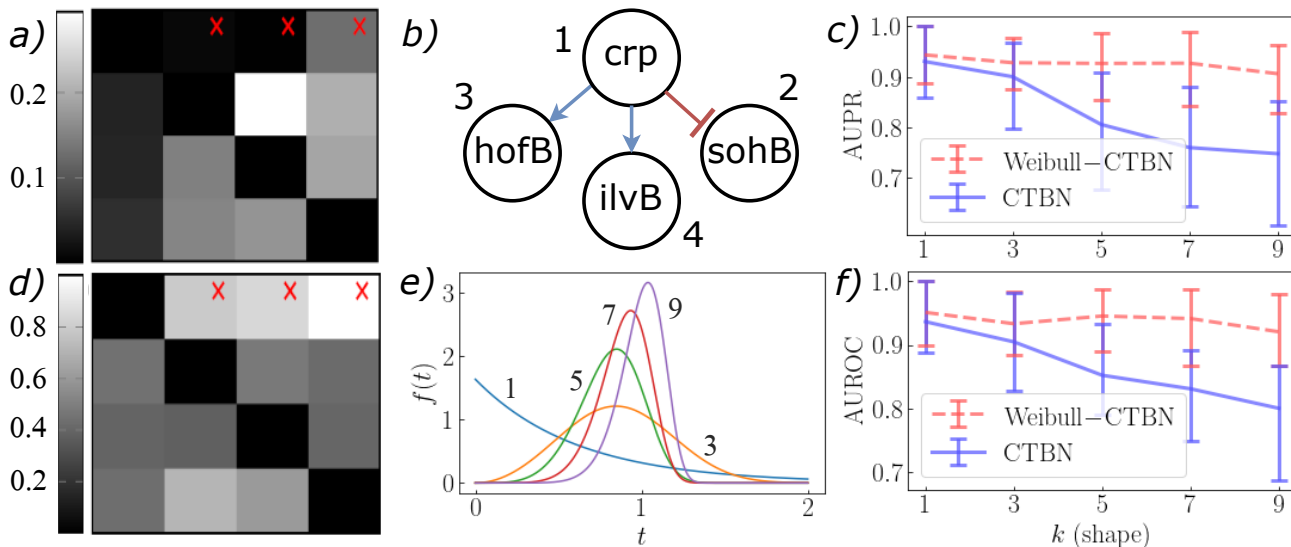
Figure 3. a) The weighted adjacency matrix for the gene regulatory network experiment obtained via structure inference from a traditional CTBN, while d) gives the result for the Gamma CTBN. The grayscales indicate the values of the posterior distribution. The traditional CTBN predicts wrong edges, while the Gamma CTBN creates an almost fully connected graph despite a Laplace prior on the edges. The high confidence edges, however, are located at the correct positions in the Gamma case. b) The GRN, from which the data has been generated. It is sourced from the gold standard for E. Coli stored in GeneNetWeaver. c) & f) The AUROC scores over the chosen shape parameters for the Weibull distributions in the structure inference scenario with controlled shape parameters. e) An illustration of the five Weibull distributions with fixed shape. The number next to each curve corresonds to its shape parameter value k. The rate parameters stayed random for each graph, process, local state and parent state combination.

Gamma CTBN, but ultimately fail to classify the true graph in both scenarios for the given amount of trajectories. The augmented CTBN, however, converged to the true graph in both scenarios.

5.1.3. EFFECTS OF SHAPE PARAMETERS

Since both Weibull and Gamma distributions contain the exponential distribution as a special case for unit shape parameters, we designed a more controlled structure inference scenario for fixed shape parameter values to study its performance impact on traditional CTBN's. Since this impact has shown to be more pronounced for Weibull distributions (see fig. 2), we chose five different fixed shape parameters k ranging from 1 to 9, increasingly distinguishing the result from an exponential distribution. We then conducted structure inference on 100 otherwise random graphs, as we did in section 5.1.2, and calculated posterior updates for 50 trajectories each. In fig. 3, we plot the median AUROC classification score along with the 0.2 and 0.8 quantiles over the chosen shape parameters for the inferred graphs. It is visible that at larger values, the scores reached by the traditional CTBN become increasingly worse compared to the ones for Weibull CTBN's. The small difference at shape $k = 1$, where all distributions are exponential, can be explained by the different prior distributions for the parameters.

Nonetheless, it is notable, that the Weibull CTBN shows no disadvantage to the traditional CTBN when confronted with traditional data. The fluctuations in the median can be accounted to numerical inaccuracies and the limited amount of samples. The equal shapes for all involved distributions, however, may also affect the scores for both the Weibull and traditional CTBN's.

5.2. Gene Regulatory Network

Finally, we show a scenario on structure inference via time-course data of gene expression through GeneNetWeaver (Schaffter et al., 2011). For this, we selected a gene regulatory network (GRN) from E. Coli consisting of 4 genes (nodes), see fig. 3 b), and simulated noise-free trajectories at maximum time resolution for a virtual time of 1000 units. We further disabled self-regulation. The exact settings can be found in Appendix B. The values in the trajectories correspond to a relative activation of the gene based on concentrations of species present in the network and therefore continuously range from 0 to 1.

As at this stage, we have not formulated our model as a latent model, we need to preprocess data. At first, we thresholded the obtained data to two levels 0 and 1 corresponding to two states per node. To account for the incompleteness,

we interpolated the observation points in the trajectories to hold their current value until the next point, at which they instantly change or not, depending on their value. After we now obtained continuous time discrete trajectories, we removed those, which exhibit less than 8 transitions. The remaining 100 trajectories were used for structure inference in the same way like explained above in the respective experiment.

We calculated the marginal probabilities of the presence of an edge using the graph posterior and arranged them in a weighted adjacency matrix. We also performed structure inference using a traditional CTBN on the data. In fig. 3 a) and d) we plot the resulting matrices for both CTBNs. We find that the high-confidence predictions by augmented CTBNs are the correct, while high-confidence predictions of traditional CTBN are incorrect for the given amount of samples.

6. Conclusion

In this work, we have presented an augmentation of CTBN's, which aims to inherit all key characteristics of its ancestor, but alleviate its restriction to simple CTMC's, which only allow exponential survival times between state changes. Inference on the model proves to be surprisingly tractable, but the introduction of real-valued clocks raises the models complexity by a significant degree. However, since computational resources are becoming more and more available, it is easier to afford inferring models bearing higher computational cost than it was in the times CTBN's were introduced. Since progress has been made in the field of approximate inference in continuous time discrete state systems with incomplete and noisy evidence recently (El-Hay et al., 2010; Rao & Teh, 2012; Linzner & Koeppl, 2018; Linzner et al., 2019), more general of such models are not longer considered hopelessly intractable. Especially since PH-distributions are highly over-parameterized and their advantages in the use in CTBN's are restricted to cases where incomplete evidence is present, we aim to present a more general approach to arbitrary survival times in CTBN's here in the light of recent progress made in the field.

Acknowledgements

References

Cassandras, C. G. and Lafortune, S. *Introduction to discrete event systems*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387333320.

Dagum, P., Galper, A., and Horvitz, E. Dynamic network models for forecasting. *Uncertainty in Artificial Intelligence*, (UAI):41–48, 1992. doi: 10.1016/b978-1-4832-8287-9.50010-4.

De, A., Valera, I., Ganguly, N., Bhattacharya, S., and Gomez-Rodriguez, M. Learning and forecasting opinion dynamics in social networks. *Advances in Neural Information Processing Systems*, (NIPS):397–405, 2016. ISSN 10495258.

D.R. Cox, D. O. *Analysis of survival data*. 1984. doi: 10.1201/9781315137438.

El-Hay, T., Cohn, I., Friedman, N., and Kupferman, R. Continuous-time belief propagation. *Proceedings of the 27th International Conference on Machine Learning*, (ICML):343–350, 2010.

Gopalratnam, K., Kautz, H., and Weld, D. S. Extending continuous-time bayesian networks. *Proceedings of the 20th National Conference on Artificial Intelligence*, (AAAI):981–986, 2005.

Hobolth, A., Siri-Jégousse, A., and Bladt, M. Phase-type distributions in population genetics. *Theoretical Population Biology*, 127:16–32, 2019. ISSN 10960325. doi: 10.1016/j.tpb.2019.02.001.

Klann, M. and Koeppl, H. Spatial simulations in systems biology: from molecules to cells. *International Journal of Molecular Sciences*, 13(6):7798–7827, 2012. ISSN 1422-0067. doi: 10.3390/ijms13067798.

Limnios, N. and Oprisan, G. *Semi-Markov processes and reliability*. Birkhäuser Basel, 01 2001. doi: 10.1007/978-1-4612-0161-8.

Linderman, S. W., Adams, R. P., and Pillow, J. W. Bayesian latent structure discovery from multi-neuron recordings. *Advances in Neural Information Processing Systems*, (NIPS):2010–2018, 2016. ISSN 10495258.

Linzner, D. and Koeppl, H. Cluster variational approximations for structure learning of continuous-time Bayesian networks from incomplete data. *Advances in Neural Information Processing Systems*, (NeurIPS):7880–7890, 2018.

Linzner, D., Schmidt, M., and Koeppl, H. Scalable structure learning of continuous-time Bayesian networks from incomplete data. *Advances in Neural Information Processing Systems*, (NeurIPS):1–11, 2019.

Liu, M., Stella, F., Hommersom, A., and Lucas, P. J. Making continuous-time Bayesian networks more flexible. *Proceedings of Machine Learning Research*, 72(PMLR): 237–248, 2018.

Maes, C., Netočný, K., and Wynants, B. Dynamical fluctuations for semi-Markov processes. *Journal of Physics A: Mathematical and Theoretical*, 42(36):365002, aug 2009. doi: 10.1088/1751-8113/42/36/365002.

Nodelman, U., Shelton, C. R., and Koller, D. Continuous time Bayesian networks. *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, (UAI): 378–387, 1995.

Nodelman, U., Shelton, C. R., and Koller, D. Learning continuous time Bayesian networks. *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, (UAI):451–458, 2003.

Nodelman, U., Shelton, C. R., and Koller, D. Expectation maximization and complex duration distributions for continuous time Bayesian networks. *Proc. Twenty-first Conference on Uncertainty in Artificial Intelligence*, (UAI):pages 421–430, 2005.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008. ISSN 00280836. doi: 10.1038/nature07140.

Pinsky, M. and Karlin, S. *An introduction to stochastic modeling: fourth edition*, pp. 1–563. Elsevier Inc, 12 2010. ISBN 9780123814166. doi: 10.1016/C2009-1-61171-0.

Rao, V. and Teh, Y. Mcmc for continuous-time discrete-state systems. *Advances in Neural Information Processing Systems*, 1(NIPS):701–709, 01 2012.

Schaffter, T., Marbach, D., and Floreano, D. GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16):2263–2270, 2011. ISSN 13674803. doi: 10.1093/bioinformatics/btr373.

Stella, F., Acerbi, E., Zelante, T., and Narang, V. Gene network inference using continuous time bayesian networks: A comparative study and application to th17 cell differentiation. *BMC Bioinformatics*, 15, 12 2014. doi: 10.1186/s12859-014-0387-x.