Infinite Continuous-time Bayesian Networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

Continuous-time Bayesian networks (CTBNs) constitute a general and powerful framework for modeling continuous-time stochastic processes on networks. We consider the case of an infinitely large CTBN. We derive a master equation, describing CTBN dynamics in the thermodynamic limit. Further we derive a set of ODEs for the moments. We use the master equation to perform approximate posterior inference for this system. Lastly, we demonstrate how to infer the degree distribution of the underlying network from data.

1 Introduction

1.1 Continuous-time Bayesian networks

We consider continuous-time Markov chains (CTMCs) $\{X(t)\}_{t\geq 0}$ taking values in a countable state-space $\mathcal{S}$. A time-homogeneous Markov chain evolves according to an intensity matrix $r : \mathcal{S}\times\mathcal{S} \to \mathbb{R}$, whose elements are denoted by $r(s, s')$, where $s, s' \in \mathcal{S}$. A continuous-time Bayesian network [**?**] is defined as an N-component process over a factorized state-space $\mathcal{S} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ evolving jointly as a CTMC. For local states $x_n, x_n' \in \mathcal{X}_n$, we will drop the states' component index n, if evident by the context and no ambiguity arises. We impose a directed graph structure $\mathcal{G} = (V, E)$, encoding the relationship among the components $V \equiv \{V_1, \dots, V_N\}$, which we refer to as nodes. These are connected via an edge set $E \subseteq V \times V$. This quantity – the structure – is what we will later learn. The instantaneous state of each component is denoted by $X_n(t)$ assuming values in $\mathcal{X}_n$, which depends only on the states of a subset of nodes, called the parent set $\mathrm{pa}(n) \equiv \{m \mid (m, n) \in E\}$. Conversely, we define the child set $\mathrm{ch}(n) \equiv \{m \mid (n, m) \in E\}$. The dynamics of a local state $X_n(t)$ are described as a Markov process conditioned on the current state of all its parents $U_n(t)$ taking values in $\mathcal{U}_n \equiv \{\mathcal{X}_m \mid m \in \mathrm{pa}(n)\}$. They can then be expressed by means of the conditional intensity matrices (CIMs) $r_n^u : \mathcal{X}_n \times \mathcal{X}_n \to \mathbb{R}$, where $u_n \equiv (u_1, \dots u_L) \in \mathcal{U}_n$ denotes the current state of the parents ($L = |\mathrm{pa}(n)|$). Specifically, we can express the probability of finding node n in state x' after some small time-step h, given that it was in state x at time t with $x, x' \in \mathcal{X}_n$ as

$$P(X_n(t+h) = x' \mid X_n(t) = x, U_n(t) = u) = \mathbf{\Delta}_{x,x'} + r_n^u(x, x')h + o(h), \tag{1}$$

where $r_n^u(x, x')$ is the matrix element of r_n^u corresponding to the transition $x \to x'$ given the parents' state $u \in \mathcal{U}_n$. It holds that $r_n^u(x, x) = -\sum_{x' \neq x} r_n^u(x, x')$. The CIMs are connected to the joint intensity matrix r of the CTMC via amalgamation – see, for example, [**?**].

2 Master equation for infinite CTBNs

In the following we want to derive a master-equation for CTBNs in the thermodynamic limit ($N \to \infty$). For this, we first consider the transition matrix of the corresponding CTMC, and show how it can be expressed in occupational numbers, that remain meaningful in the thermodynamic limit. After this we use this exact transition matrix, to derive a master-equation for infinite tree-like graphs.

Submitted to 32nd Conference on Neural Information Processing Systems (NIPS 2018). Do not distribute.

2.1 Transition matrix for an infinite CTBN

Let us denote $\mathcal{G}$ as the set of possible local graphs in a network. Let $d : V \to \mathbb{N}$ be a function that maps the index of every node to its number of ingoing edges. We will refer to this number as the *degree* of the node. Assuming a homogeneous system (the conditional transition matrices depend only on the degree, the nodes state x_n and its parents state u_n), we can then write the transition matrix of the factored CTMC as

$$\lim_{h\to 0} P(X(t+h) = s' \mid X(t) = s) = \lim_{h\to 0} \prod_{n=1}^{N} P_{d(n)}(X_n(t+h) = y_n \mid X_n(t) = x_n, U_n(t) = u_n)$$

In the following we will define the shorthand $P_k^h(y \mid x, u) \equiv P_{d(n)=k}(X_n(t+h) = y_n \mid X_n(t) = x_n, U_n(t) = u_n)$. Because of the homogueniety, we can re-oder the product

$$\lim_{h\to 0} P(s' \mid s) = \lim_{h\to 0} \prod_{k=0}^{N} \prod_{u\in\mathcal{U}_k} \prod_{x\in\mathcal{X}} P_k(y \mid x, u)^{n_k^u(x)},$$

where k is the degree, u the parents-state (which is limited by the nodes degree) and z the state of the node. It holds that $\sum_{k,u,z} n_k^u(z) = N$. By inserting (1) and using Bernoullis' theroem we get

$$\lim_{h\to 0} P(s' \mid s) = \lim_{h\to 0} \prod_{k=0}^{N} \prod_{u\in\mathcal{U}_k} \prod_{x\in\mathcal{X}} \left\{ \mathbf{\Delta}_{x,y} \sum_{l=0}^{n_k^c(x)-1} \begin{pmatrix} n_k^c(x) \\ l \end{pmatrix} (hr_k^u(x,y))^l + (hr_k^u(x,y))^{n_k^u(x)} \right\}.$$

We make the assumption on the graph that $n_k^c(x) \to \infty$ in the limit $N \to \infty$. Thus, we can and employ Stirlings' approximation $\begin{pmatrix} n_k^c(x) \\ l \end{pmatrix} \to \frac{(n_k^c(x))^l}{l!}$ and arrive at

$$\lim_{N\to\infty} \lim_{h\to 0} P(s' \mid s) = \lim_{N\to\infty} \lim_{h\to 0} \prod_{k=0}^{N} \prod_{u\in\mathcal{U}_k} \prod_{x\in\mathcal{X}} \left\{ \mathbf{\Delta}_{x,y} \exp\left\{hn_k^c(x) r_k^c(x,y)\right\} + (hr_k^c(x,y))^{n_k^c(x)} \right\}.$$

If $s' = s$, then it follows in the limit $N \to \infty$

$$\lim_{N\to\infty} \lim_{h\to 0} P(s \mid s) = \lim_{N\to\infty} \lim_{h\to 0} 1 + hNR(s,s) + o(h),$$

with the expected *global* rate $R(s,s) \equiv \sum_{k,u,x} p_k^u(x) r_k^u(x,x)$, with $p_k^u(x) \equiv \frac{1}{N} n_k^u(x)$ being the *fraction* of nodes of type $\{k, u, x\}$. If $y \neq x$, we use that, because of the asynchronous update constraint on CTBNs, only one node (which we will denote by $*$) can change states at a time for $h \to 0$, thus

$$\begin{aligned} \lim_{N\to\infty} \lim_{h\to 0} P(s' \mid s) &= \lim_{N\to\infty} \lim_{h\to 0} \frac{hr^*(x,y)}{1 + hr^*(x,x)} \exp\left\{1 + hNR(s,s)\right\} \\ &= \lim_{h\to 0} hr^*(x,y), \end{aligned}$$

as $o(h)$ therms vanish in the limit. We find, that the transition matrix of an infinite CTBN only depends on the occupational numbers of state s, the degree, parents'-state and the state of the node undergoing a transition.

2.2 Lumping: Transitions in many-body basis

In the following, we want to transform the transition matrix to an equivalent basis. For this, it is sufficient to determine the matrix elements in this new basis. In the following we want to restrict ourselves to binary, dynamics $\mathcal{X}_n = \{0, 1\}$. In order to perform the transformation, we have to define mappings. Further, we replace u by its summary $m = \sum_u$ the number of parents in state 1, assuming that $r_k^u(x,y) = r_k^m(x,y)$. For a binary system, it is also sufficient to denote the exist state of the rate $r_k^m(x) \equiv r_k^m(x, \bar{x})$. We define functions $\phi : V \times \mathcal{S} \to \mathbb{N}$ mapping a node index, to the number of parents in state 1 with respect to the global state s. Further, we the function $\psi : V \times \mathcal{S} \to \mathcal{X}$ maps a node index and the global state s, to its local state with respect to s. We define the *flip-operator* $F : V \times \mathcal{S} \to \mathcal{S}$. $F(i, s)$ flips the i'th element of s, e.g. $F(2, (0, 0, 1, 0)^T) = (0, 1, 1, 0)$.

Definition 1. *We define a non-invertible mapping $\mathcal{M} : \mathcal{S} \to \mathbb{N}^{K\times K\times|\mathcal{X}|}$, where K is the maximum degree occuring in the infinite graph, corresponding to a* lumping *of the state space. The mapping $\mathbf{M}(\mathbf{s})$ is constructed by counting the $n_k^m(x)$ of s. The occupation numbers of state s are given as elements $n_k^m(x) = [\mathbf{M}(\mathbf{s})]_{k,m,x}$. The image of $\mathcal{M}$ is a vector space.*

Example 1. *We can represent the mapping as state vectors of the type $\mathbf{M}(\mathbf{s}) = \left(n_0^0(x), \dots, n_k^m(0), \dots\right)^T$. As only local transitions are possible, a valid exit state is $\mathbf{M}(\mathbf{s}') = \left(n_0^0(0), \dots, n_k^m(0)-1, \dots, n_k^m(1)+1, \dots, n_{k'}^{m+1}(0)+1, \dots, n_{k'}^m(0)-1, \dots\right)^T$. In this case, a node of degree k switched state $0 \to 1$. This node had only one child in state 0 of degree k' with parents state m.*

For the following analysis, we define the set of possible neighborhoods $G_i \equiv \{k \mid k = deg(j), j \in \mathrm{par(i)}\}_{i\in V}$ and $G \equiv \bigcap_{i=1}^N G_i$. In order to consider the general case we can write

$$R(\mathbf{M}(\mathbf{s}') \mid \mathbf{M}(\mathbf{s})) = \begin{cases} hr_{d(i)}^{\phi(i,s)}(\psi(i,s), \psi(i,s')) & , \mathbf{M}(\mathbf{s}') = \mathbf{M}(\mathbf{s}) \\ & +e_{d(i),\phi(i,s'),\psi(i,s')} - e_{d(i),\phi(i,s),\psi(i,s)} + e_{g_{i,s}} \\ hNR(s,s) & , s' = s \\ 0 & , \text{else} \end{cases}$$

for all $i \in V$ with the *neighborhood* $g_{i,s} \equiv \sum_{j\in \mathrm{ch}(i)} \left(e_{d(j),\phi(j,s)+(-1)^{\psi(i,s)},\psi(j,s)} - e_{d(j),\phi(j,s),\psi(j,s)}\right)$ and unit-vectors $e_{x,y,z} \in \{0,1\}^3$. Lastly, we define a projection operator

Definition 2. *Let $v \subseteq V$ be a subset of the set of vertices. We define a projection operator $\mathbb{P}_v : \mathcal{S} \to \{\mathcal{X}_i \mid i \in v\}$, that projects a state $s \in \mathcal{S}$ on a subset $\{\mathcal{X}_i \mid i \in v\} \subseteq \mathcal{S}$.*

The probability to find a system in state $m = \{s \in \mathcal{S} \mid \mathbf{M}(\mathbf{s}) = m\}$ is $R(m' \mid m) = \sum_s \delta(\mathbf{M}(\mathbf{s}) = m)R(m' \mid \mathbf{M}(\mathbf{s}))$

Remark. which we will extensively use throughout the manuscript.

2.3 (Approximate) Master equation for lumped CTBNs

We consider all possible transitions ending in state m

$$\begin{aligned}\partial_t P_t(m) &= \sum_{k,m,x}\sum_{g\in G_k} R\left(m \mid m - \Delta^g_{k,m,x}\right) P_t\left(m - \Delta^g_{k,m,x}\right) - R\left(m + \Delta^g_{k,m,x} \mid m\right) P_t(m), \\ &= \sum_{k,m,x} r_{k,m}(x)\left\{\sum_{g\in G_k} n^g_{k,m,x} P_t\left(m - \Delta^g_{k,m,x}\right) - (n_{k,m,x} + [\sum_{g\in G_k}\Delta^g_{k,m,x}]_{k,m,x})P_t(m)\right\},\end{aligned}$$

with

$$\begin{aligned}\Delta^g_{k,m,x} &\equiv e^g_{k,m,\bar{x}} - e^g_{k,m,x} + \sum_{\{k',m',x'\}\in g}\sum_{g'\in G_{k'}} n^{g'\cdot g}_{k',m',x'}[e^{g'}_{k',m'+(-1)^x,x'} - e^{g'}_{k',m',x'}], \\ &\approx e^g_{k,m,\bar{x}} - e^g_{k,m,x} + \frac{k}{N}\sum_{\{k',m',x'\}\in g}\sum_{g'\in G_{k'}} p(k' \mid k) n^{g'}_{k',m',x'}[e^{g'}_{k',m'+(-1)^x,x'} - e^{g'}_{k',m',x'}],\end{aligned}$$

where we assumed that all nodes of type $\{g, k, m, x\}$ are statistically equivalent.

2.4 Van Kampen Expansion

We make the standard Van Kampen expansion $n^g_{k,m,x} = N\phi^g_{k,m,x}(t) + \sqrt{N}\eta^g_{k,m,x}$, and transform variables $P_t(m) = N^{-1/2}\Pi_t(\eta)$. The derivative becomes $\partial_t P_t(m) = N^{-1/2}\partial_t\Pi_t(\eta) - \partial_\eta\Pi_t(\eta)\partial_t\phi(t)$ then

$$\partial_t\Pi(\eta;t) - N^{1/2}\sum_{k,m,x}\sum_{g\in G_k}\partial_t\phi^g_{k,m,x}(t)\partial_{\eta^g_{k,m,x}}\Pi(\eta;t) =$$

$$\sum_{k,m,x} r_{k,m}(x)\sum_{g\in G_k}\left\{n^g_{kmx}N^{-1/2}\partial_{\mathbf{\Delta}^g_{kmx}} - [\sum_{g\in G_k}\mathbf{\Delta}^g_{kmx}]_{kmx}\right\}\Pi_t(\eta),$$

$$\partial_t \Pi(\eta;t) - N^{1/2} \sum_{k,m,x} \sum_{g \in G_k} \partial_t \phi^g_{k,m,x}(t) \partial_{\eta^g_{k,m,x}} \Pi(\eta;t) =$$
$$\sum_{k,m,x} r_{k,m}(x) \sum_{g \in G_k} \left\{ \left(N^{1/2} \phi^g_{k,m,x}(t) + \eta^g_{k,m,x} \right) \partial_{\mathbf{\Delta}^g_{kmx}} - [\sum_{g \in G_k} \mathbf{\Delta}^g_{kmx}]_{kmx} \right\} \Pi_t(\eta),$$

$$\partial_t \Pi(\eta;t) - N^{1/2} \sum_{k,m,x} \sum_{g \in G_k} \partial_t \phi^g_{k,m,x}(t) \partial_{\eta^g_{k,m,x}} \Pi(\eta;t) =$$
$$\sum_{k,m,x} r_{k,m}(x) \sum_{g \in G_k} \left\{ \left(N^{1/2} \phi^g_{k,m,x}(t) + \eta^g_{k,m,x} \right) \partial_{\mathbf{\Delta}^g_{kmx}} - kp(k \mid k) \phi_{k,m+(-1)^x,x} + 1 \right\} \Pi_t(\eta),$$

2.4.1 Van Kampen $O(N^{1/2})$

$$\begin{aligned}
&\sum_{k,m,x} \sum_{g \in G_k} \partial_t \phi^g_{k,m,x}(t) \partial_{\eta^g_{k,m,x}} \Pi(\eta;t) = \sum_{k,m,x} r_{k,m}(x) \sum_{g \in G_k} \phi^g_{k,m,x}(t) \partial_{\mathbf{\Delta}^g_{kmx}} \Pi_t(\eta), \\
&= \sum_{k,m,x} r_{k,m}(x) \sum_{g \in G_k} \phi^g_{k,m,x}(t) \Big\{ \partial^g_{\eta_{k,m,\bar{x}}} - \partial^g_{\eta_{k,m,x}} \\
&+ k \sum_{\{k',m',x'\} \in g} \sum_{g' \in G_{k'}} p(k' \mid k) \left[\phi^{g'}_{k',m',x'} + N^{-1/2} \eta^{g'}_{k',m',x'} \right] \left[\partial_{\eta^{g'}_{k',m'+(-1)^x,x'}} - \partial_{\eta^{g'}_{k',m',x'}} \right] \Big\} \Pi_t(\eta),
\end{aligned}$$

We assume a naive mean field perspective $\phi(t) = \prod_{g,k,m,x} \phi^g_{k,m,x}(t)$. We can then sort the equations per $\partial_{\eta^g_{k,m,x}} \Pi_t(\eta)$ and get

$$\begin{aligned}
\partial_t \phi^g_{k,m,x}(t) =& r_{k,m}(x) \phi^g_{k,m,\bar{x}}(t) - r_{k,m}(\bar{x}) \phi^g_{k,m,x}(t) \\
&+ k(\phi^g_{k,m+(-1)^x,x}(t) - \phi^g_{k,m,x}(t)) \sum_{\{k',m',x'\} \in g} \sum_{g' \in G_{k'}} r_{k',m'}(x') p(k' \mid k) \phi^{g'}_{k',m',x'}(t)
\end{aligned}$$