Technical Report: CS-03-02

Monte-Carlo Approximations to Continuous-time Semi-Markov Processes

Milos Hauskrecht
Department of Computer Science
University of Pittsburgh
milos@cs.pitt.edu

Abstract

Continuous-time semi-Markov processes (SMP) represent a very useful extension of Markov process models to dynamic systems whose behaviors are more naturally described in terms of time profiles. The semi-Markov model, however, comes with a limitation: individual transitions in the model can occur at different times which means that for a specific time interval T the system modeled as an SMP can go through exponentially many transitions. This affects adversely the worst-case performance of Monte Carlo methods designed to support various inference and decision tasks. In this work, we propose and analyze modifications of Monte Carlo methods that rely on simulation trajectories of a limited length N. We prove that for any practical time models the probability of not reaching the target time T by a trajectory of length N decays exponentially fast. Using this result we analyze extensions of some of the Monte Carlo inference methods to semi-Markov settings and derive sample complexity bounds for additive $\epsilon\delta$ estimates of desired quantities.

1 Introduction

Markov processes (MPs) form the foundations of work on modeling stochastic dynamic systems in Artificial Intelligence. The AI research has traditionally focused on discrete-time Markov process models, their efficient representations and

inference solutions. The Markov property enables the time decomposition of the MP model and the same time-invariant behavior in every transition. This greatly simplifies the analysis of the model and associated inferences. However, Markov models are not fit well to represent behaviors that are more naturally described in terms of time profiles. As has been pointed out by some researchers such processes are common in many application areas including medicine [12], transportations [2], robotics [13, 7], and others [5, 3].

Semi-Markov processes (SMPs) [9, 10, 17] alleviate the time-invariance problem and extend Markov processes to time-dependent transitions. In continuous-time SMPs state transitions can occur at any time. Similarly to discrete-time Markov processes, continuous-time SMPs allow us to decompose the process to independent time segments bounded by state transitions. However, a complication is that durations of time segments can vary. As a result, inferences in semi-Markov processes must consider cases in which a large number of transitions occurs within a time segment T. This problem can affect the worst case performance of Monte Carlo (MC) methods that perform the inference by simulating many different state trajectories over a specific time segment.

To address the problem of 'long' trajectories we propose and analyze a simple modification of Monte Carlo sampling methods for SMPs. The idea of the modification is to use only trajectories with up to N transitions, all other (longer) trajectories are replaced with a random guess. We show that for **any practical time density model** (a density model with a finite mean) the probability of generating trajectories of length more than N decays exponentially fast. This is the key result that allows us to modify many Monte Carlo algorithms developed for discrete-time models and derive sample complexity bounds for additive $\epsilon\delta$ approximations. We illustrate the results on the analysis of algorithms for state prediction, but we expect a similar transfer to apply also to decision-making problems.

In the following we first introduce continuous-time semi-Markov processes and discuss their compact representations. Next we consider the problem of state prediction and propose a simple Monte Carlo algorithm that works with truncated state-time trajectories. We prove the key theorem of the paper – the exponential decay theorem for truncated simulation trajectories and illustrate it on semi-Markov models with Gamma time densities. Finally, we use the theorem to analyze and derive sample complexity bounds of some MC state-prediction algorithms.

2 Continuous-time Semi-Markov Process

A *semi-Markov process* (SMP) [9, 10, 1] extends discrete-time Markov process models to time-dependent stochastic behaviors. A semi-Markov process is like a Markov process except that transition probabilities depend on the amount of time elapsed since the last change in the state. The Markov property holds as a consequence of the reset property; once a change occurs (and we know the state) the future becomes independent of the past.

More formally, a (finite-state) semi-Markov process (SMP) is defined by a tuple (S, Φ, s_0) where: S is a finite set of states and Φ is a stochastic transition model such that $\phi(s', t|s)$ represents the probability of a transition (s, s') occurring in time interval $[0, t]$; and s_0 is the initial state. In most cases, the transition model ϕ of an SMP is not provided directly. Instead, transition probabilities are defined as:

$$\phi(s', t|s) = \int_0^t P(s'|s)p(t'|s, s')dt' \quad = \quad P(s'|s) \int_0^t p(t'|s, s')dt'$$

where $P(s'|s)$ is the probability of a transition (s, s'), as used in a Markov process, and $p(t|s, s')$ models the distribution of stochastic transition times. A semi-Markov model approximates the dynamics of a continuous-time system by decomposing the process along times in which the state change occurs and its time model reflects the fluctuation of these changes in time.

Compact Parameterizations. In general, a semi-Markov process permits models of arbitrary complexity. However, for practical purposes we are interested in models with efficient parameterizations. We focus on two aspects of parameterization: (1) parameterization of transition time densities $p(t|s, s')$; (2) parameterizations of state transitions $p(s'|s)$.

2.1 Compact Parameterizations of Time Densities

Time dependencies in SMPs are modeled through time densities $p(t|s, s')$ reflecting the fluctuation of transition times during the transition between states s and s'. Time models are defined over interval $[0, \infty]$. For the sake of representational efficiency we are interested primarily in densities with compact parameterizations. A good time density model candidates are: Gamma, lognormal or a Gaussian density rectified to $[0, \infty]$ range. Figure 1 gives examples of time models based on *Gamma distribution* and its mixtures. Gamma distribution belongs to the family of

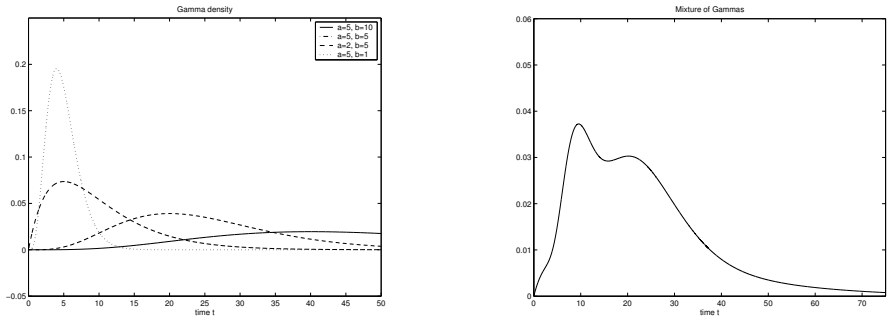

Figure 1: (a) Gamma density for 4 different parameter settings. (b) Mixture of 4 Gammas.

exponential distributions. The density function for Gamma distribution is:

$$f(t|a,b) = \frac{1}{b^a \Gamma(a)} t^{a-1} e^{-\frac{t}{b}},$$

where a is a shape and b is a scale parameter and $\Gamma(a)$ is a Gamma function. The mixture of Gammas model is a convex combination of multiple Gamma models: $p(t|s,s') = \sum_{i=1}^{K} w_i p_i(t|s,s')$, where $0 \leq w_i \leq 1$ are weight parameters such that $\sum_{i=1}^{K} w_i = 1$ and $p_i(t|s,s')$ is a Gamma model in the mixture. The advantage of the mixture model is that it gives us more flexibility in representing more complex time profiles.

2.2 Factored-state Semi-Markov Models

Much of the recent research work in modeling Markov processes focuses structured representations based on *dynamic belief networks (DBNs)* [6, 11]. Such models allows us to represent processes over large state spaces more efficiently. In particular, the state of the process is factored into a set of state variables $\mathcal{S} = \{S_1, S_2, \ldots, S_k\}$ and transition probabilities $p(s'|s)$, decompose to 'local' transition probabilities $p(s'_i|pa(s'_i))$, such that $pa(s'_i)$ denote state variables the state variable S_i depends on. Our goal is to extend factored representations to semi-Markov processes. The key is to model appropriately time dependent interactions between an arbitrary state variable and all state variables it depends on. In particular, it is necessary to define a clock for each state variable S_i and rules for its resets. Nodelman et al [16]

proposed one such model for a special class of homegeneous Markov processes. We extend the model to a more general semi-Markov settings.

Let $C(S_i)$ be a set of state variables the variable S_i depends on. Our transition model for factored semi-Markov processes is defined by:

- **Assumption 1:** only one state variable can change at any instance of time;
- **Assumption 2:** a change in values of S_i or $C(S_i)$ resets the time for S_i to 0;
- The **transition model** for S_i and fixed values of its conditioning set $c(S_i)$ is a semi-Markov process model defined in terms of two probabilities:
  - $P(s'_i|s_i, c(S_i))$ - the probability of a transition (s_i, s'_i) under the fixed set of values $c(S_i)$ of the conditioning set.
  - $p(t|s_i, s'_i, c(S_i))$ - time density for $(s_i s'_i)$ under $c(S_i)$.

Basically, the time changes in S_i are conditionally independent of the rest of the state variables in $\mathcal{S}$ given $C(S_i)$. In addition, the only events affecting the time profile resets of S_i are changes in conditioning variables and in S_i itself.

3 MC Algorithms for State Predictions in SMPs

Probabilistic inference in context of stochastic dynamic models covers predictions, diagnosis, and other probabilistic queries. In this work, we focus on the state prediction problem where we are interested in computing the probability $P(s_T = v|s_0)$ of a future state of the system in time T given the initial state s_0 or given the distribution of initial states. In general, no closed form solution for computing the probability of $P(s_T = v|s_0)$ or the distribution $P(s_T|s_0)$ for an arbitrary time-model exists. Exceptions are special cases, such as homogeneous Markov processes [1, 16] with exponentially distributed transition times.

3.1 Basic Monte Carlo Inference Algorithm

Monte-Carlo (MC) approximations are a natural choice when closed form inference solutions do not exist. The basic Monte-Carlo algorithm for state value predictions is simple:

1. **generate** M samples of states at time T by sampling sequentially transitions (states and times) of a semi-Markov model (starting from the initial state s_0);

2. **estimate** the target probability $P(s_T = v|s_0)$ of a state s_T using the frequency of occurrences of value v in M samples.

Complexity analysis. The total number of simulation steps depends on the number of samples M and lengths of simulation trajectories in every sample. Assuming the same time model for every transition, the dependence between the error ϵ, confidence parameter δ and the number of samples M is captured by the following (rather trivial) theorem.

Theorem 1 *Let $\mu_M(v)$ be the sample average approximation of $P(s_T = v|s_0)$. Then: $P\left(|\mu_M(v) - E(s_T = v|s_0)| \geq \epsilon\right) \leq 2e^{-\lceil 2\epsilon^2 M \rceil}$. The number of sample trajectories that guarantees the $\epsilon\delta$ approximation is $M \geq \frac{1}{2\epsilon^2} \ln \frac{2}{\delta}$.*

Proof. Direct application of Hoeffding's inequality [8, 15]. ■

To fully analyze the complexity of the Monte Carlo algorithm we need to consider the number of state transitions in every simulation trajectory reaching time T. The problem is that transitions in a semi-Markov model may occur at different times and trajectories with (exponentially) many transitions (in terms of states) may be occasionally generated. This affects negatively the worst case complexity of the basic MC solution. However, we note that in the average-case analysis the expected number of transitions in a trajectory depends only on T and the mean of the time model and equals $E(N) = \lceil \frac{T}{E(t)} \rceil$.

3.2 Modified MC Algorithm

The problem with the basic Monte Carlo algorithm is the worst case complexity since some trajectories may become very long. To remedy the problem we propose a modification of the MC algorithm that simulates trajectories for at most N steps. The algorithm works as follows:

- **generate** M samples of states at time T by sampling sequentially transitions (states and times) of a semi-Markov model for at most N steps, **guess randomly** the result for samples not reaching time T;
- **estimate** the target probability $P(s_T = v|s_0)$ of a state $s_T = v$ using M samples.

We will show that for an appropriate choice of N, that is logarithmic in $1/\epsilon$, we obtain an additive $\epsilon\delta$ approximation that is very similar in terms of the sample

complexity of M to the basic MC algorithm. The crucial thing is that the result generalizes to any practical transition time model with a finite mean or combinations of such models.

Probability of Incomplete Simulation Trajectories

To analyze the complexity of the modified MC algorithm we first prove the key result on probabilities of incomplete trajectories, that is, simulation trajectories that were not able to reach time T in N steps. We show that for a sufficiently large N that depends only on the characteristics of the time model density, the probability of obtaining incomplete trajectories decreases exponentially fast in N. The result holds for an arbitrary time density model with a finite mean. To simplify the initial analysis we first consider the case in which every transition in the model follows the same time density.

Theorem 2 *Let $S_N = t_1 + t_2 + \cdots t_N$ is a sum of independent random variables representing transition times of an arbitrary time density model t with a finite mean. Then for $N > T/E(t)$ the probability of S_N not reaching time T, $p_N = P(S_N < T)$, decays exponentially fast in the number of steps N, in particular,*

$$p_N = P(S_N < T) \leq A(T, \beta) e^{-NC(\beta)}$$

where $A(T, \beta) > 0, C(\beta) > 0$ are constants that depend only on the time model used, the target time T, and a variational parameter $\beta > 0$.

Proof. To bound the difference between S_N and T we build upon Chernoff's exponential bound [4] (see also [14]). Let S be a real-valued random variable with a finite mean $E(S)$. Then for any value $\beta > 0$, and $\epsilon \geq E(S)$, it holds: $P(S \geq \epsilon) \leq e^{-\beta\epsilon} E(e^{\beta S})$. If S_N is a sum of N independent random variables X_i we can rewrite the bounding expression as:

$$P(S_N \geq \epsilon) \leq e^{-\epsilon\beta} E(e^{\beta S_N}) = e^{-\beta\epsilon} \prod_{i=1}^{N} E(e^{\beta X_i}).$$

We want to bound the probability $p_N = P(S_N < T)$ of not reaching time T in N steps. First, p_N can be bounded as $p_N \leq P(E(S_N) + S_N \leq E(S_N) + T) = P(E(S_N) - S_N \geq E(S_N) - T)$. Assuming that $E(S_N) - T \geq 0$ holds , Chernoff's exponential bound [4] applies and p_N can be bounded as:

$$p_N \leq P(E(S_N) - S_N \geq E(S_N) - T) \leq e^{-\beta(E(S_N)-T)} E(e^{\beta(E(S_N)-S_N)}) = e^{\beta T} \left[E(e^{-\beta t})\right]^N .$$

The expectation is over time t. Since $e^{-\beta t} < 1$ holds for any $t > 0$ and $\beta > 0$, the expectation $E(e^{-\beta t})$ is bounded strictly from above by the distribution function:

$$\int_0^\infty e^{-\beta t} p(t|\theta) dt < \int_0^\infty p(t|\theta) dt = 1.$$

Thus, there exists a constant $1/K(\beta)$ such that $E(e^{-\beta t}) \leq 1/K(\beta) < 1$. Substituting the bound into the bound for p_N we get:

$$p_N \leq e^{\beta T - N \log K(\beta)}.$$

As $K(\beta) > 1$, the probability $p_N = P(S_N < T)$ can be bounded by $p_N = P(S_N < T) \leq A(T, \beta) e^{-NC(\beta)}$ with $A(T, \beta) > 0, C(\beta) = \log K(\beta) > 0$. Thus, given $N \geq T/E(t)$ the probability of p_N decays exponentially in N. ■

The exponential decay result is critical for the analysis of Monte Carlo algorithms for continuous-time semi-Markov models. Constants determining the decay rate are density specific and need to be derived. Take, for example, **Gamma time model**. The mean of $Gamma(t|a, b)$ is $E(t) = ab$ and $E(e^{-\beta t}) = (1 + \beta b)^{-a}$. This leads to a bound:

$$p_N \leq e^{\beta T - aN \ln(1+\beta b)} = e^{\beta T} e^{-Na \ln(1+\beta b)}.$$

Figure 2 shows exponential bounds for a Gamma distribution and three different values of β and compares them to its empirical p_N. The best value of β, found by differentiating the bound and setting the result to 0, is $\beta = Na/T - 1/b$.

Multiple different time models. The assumption used in the proof of Theorem 2 was that the transition time model is fixed for every transition in the SMP. To obtain the bound on p_N for an SMP with many different time models we can substitute all model specific quantities in the theorem with their worst case values from among all time models. Assuming that there are L different transition time models t_i we can derive, for $N > T/[\min_{i=1,\cdots,L} E(t_i)]$:

$$p_N \leq A(T, \beta) e^{-N \log K^*(\beta)},$$

where $1/K^*(\beta) \geq \max_{i=1,\cdots,L} E_{t_i}(e^{-\beta t_i})$, and $A(T, \beta) = e^{\beta T} > 0$.

Complexity of the Modified MC Algorithm

Using theorem 2 we can to derive the complexity of the modified MC algorithm introduced in Section 3.2. Once again the result applies to any time density with a finite mean.

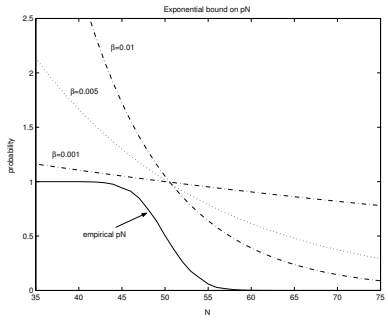

Figure 2: Gamma time model example with ($a = 5, b = 2$) and $T = 500$. An empirical estimate of p_N based on 10000 trajectories and bounds on p_N for 3 values of β are shown.

Theorem 3 *Let $\mathcal{A}$ be a modified MC algorithm that estimates the probability $P(s_T = v|s_0)$ using the sample average $\mu^T_{N,M}(v)$ based on M truncated simulation trajectories of the maximum length N. Then, for a fixed time density model t, $\mu^T_{N,M}(v)$ gives an $\epsilon\delta$ approximation $P(|\mu^T_{N,M}(v) - E(s_T = v|s_0)| \geq \epsilon) \leq \delta$ when:*

$$N \geq \max\left\{\frac{T}{E(t)}; \frac{1}{C}\left[\log\left(\frac{2}{\epsilon}\right) + T\right]\right\} \qquad \textit{and} \qquad M \geq \frac{2}{\epsilon^2}\ln\left(\frac{2}{\delta}\right),$$

where $C > 0$ is a constant specific to the time model t.

Proof.

$$\begin{aligned} &P(|\mu^T_{N,M}(v) - E(s_T = v|s_0)| \geq \epsilon) = \\ &\quad = \quad P(|\mu^T_{N,M} - E(s^N_T = v|s_0, N)| + |E(s^N_T = v|s_0, N) - E(s_T = v|s_0)| \geq \epsilon) \\ &\quad = \quad P(|\mu^T_{N,M}(v) - E(s^N_T = v|s_0, N)| \geq \epsilon - p_N) \leq 2e^{-\left[2(\epsilon - p_N)^2 M\right]}. \end{aligned}$$

Last step follows from Hoeffding's inequality. Using the bound on p_N (from theorem 2) with $\beta = 1$ to assure that p_N accounts for at most half of the error we get: $p_N \leq e^{T-NC} \leq \frac{\epsilon}{2}$. Substituting p_N and expressing both M and N we obtain:

$$N \geq \frac{1}{C}\left[\log\left(\frac{2}{\epsilon}\right) + T\right] \qquad \text{and} \qquad M \geq \frac{2}{\epsilon^2}\ln\left(\frac{2}{\delta}\right). \qquad \blacksquare$$

We note that the sample complexity result in theorem 3 is not the tightest possible and can be improved, for example, by optimizing the variational parameter

β. However, the theorem illustrates clearly two very important points. First, the 'worst-case' sample complexity result for the modified MC is comparable, in terms of M, to the 'expected' sample complexity of the basic MC algorithm (see Theorem 1). Second, the proof and the result show that the sample complexity bounds for N and M are coupled only through error parameter ϵ. Thus, using a fixed proportion of ϵ to cover p_N, N is able to fully absorb the effect of continuous time at a very modest (logarithmic in $1/\epsilon$) expense. This separation is very convenient for extending MC algorithms developed for discrete-time Markov processes into continuous-time semi-Markov settings and their subsequent sample complexity analysis.

MC Algorithm for Factored SMPs

The modified MC algorithm and its analysis for flat-state SMPs (Theorem 3) can be easily extended to factored settings. In terms of complexity bounds only N will change.

Trajectories in factored SMPs are defined by k state variable subprocesses and each of these trajectories must reach time T to make a simulation run successful. Things are further complicated by dependences among state variables. An occurrence of a transition affects all the variables that are conditioned on it and resets their clock. The reset effect may cascade through multiple dependency relations. To analyze MC algorithms for factored SMPs we consider two extreme cases:

• **Fully independent semi-Markov subprocesses.** In this case each state variable is independent of other state variables and we have k independent subprocesses. The probability of not reaching the target time T with at least one of the subprocesses can be bound through the union (Bonferoni) bound, that is: $p_N \leq ke^{T-NC}$. Then, using the same assumptions as in theorem 3, the number of steps N for every trajectory necessary to assure $\epsilon\delta$ approximation is: $N \geq \frac{1}{C}\left[\log\left(\frac{2k}{\epsilon}\right) + T\right]$. Note that M stays unchanged.

• **Fully dependent semi-Markov subprocesses.** In this case every state variable process is affected by changes in all other state variables. Assuming that $t^1, t^2, \cdots, t^K$ are random variables representing transitions times for state variables $S_1, S_2, \cdots, S_K$, the time of occurrence of a transition is $t \sim \min_{j=1\cdots K} t^j$. If the time model used for each state variable S_j is the same and follows the density $p'(t)$, the min statistic is: $p(t) \sim K(1 - F'(t))^{(K-1)}p'(t)$, where $F'(t)$ is the distribution function of $p'(t)$. Nevertheless, Theorem 3 applies directly to any time model t. But then, M stays unchanged, and the only difference is the density

specific constant C affecting the bound on N.

Extensions of the Modified MC Method

A similar disassociation pattern between M and N also transfers to extensions of the modified MC algorithm to more complex inference tasks. An example is an MC algorithm approximating the distribution $p(\mathbf{s}_T|s_0)$ of a state at time T with its empirical distribution $\mu^T_{M,N}$ based on truncated simulation trajectories. In such a case, the lower bound on the length of simulation trajectories N for an accuracy ϵ under the $L1$ norm becomes:

$$N \geq \frac{1}{C}\left[\log\left(\frac{2|S|}{\epsilon}\right) + T\right],$$

while the sample complexity bound for the number of samples M differs from the discrete-time case by the 'standard' $\epsilon \longrightarrow \epsilon/2$ correction.

4 Conclusions

Continuous-time semi-Markov processes allows us to model stochastic systems with continuous time-dependent transition profiles. This property is crucial for modeling many applications domains [12, 2, 3]. This advantage, however, comes with trade-offs: (1) time-dependencies in the semi-Markov process model must be explicitly represented which itself may translate to the increase in the complexity of the stochastic model (2) closed-form inference solutions are typically not available. To address these problems we have presented and analyzed (1) compact parameterizations of semi-Markov processes based on parametric time models and factorizations and (2) MC inference algorithms. We showed that time and simulations over time for MC state predictions do not prevent efficient worst-case inferences and a simple modification of the MC algorithm successfully resolves the concern of exponentially long simulation trajectories for any practical time model. We gave the detailed analysis of the modified MC solution and discussed its extensions to more complex inferences and models. We expect that the modification and its analysis will transfer in a rather straightforward way also to MC methods for semi-Markov decision processes.

References

[1] Dimitri P. Bertsekas. *Dynamic programming and optimal control*. Athena Scientific, 1995.

[2] Justin A. Boyan and Michael L. Littman. Exact solutions to time-dependent MDPs. In *NIPS*, pages 1026–1032, 2000.

[3] John Bresina, Richard Dearden, Nicolas Meuleau, Sailesh Ramakrishnan, David Smith, and Ric h Washington. Planning under continuous time and resource uncertainty. In *Proceedings of the Eighteen Conference on Uncertainty in Artificial Intelligence*, pages 77–85, 2002.

[4] H. Chernoff. A measure of the asymptotic efficiency of tests of a hypothesis based on the sum of observations. *Annals of Mathmematical Statistics*, 23:493–507, 1952.

[5] Robert H. Crites and Andrew G. Barto. Improving elevator performance using reinforcement learning. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 1017–1023. The MIT Press, 1996.

[6] Thomas Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5:142–150, 1989.

[7] Mohammad Ghavamzadeh and Sridhar Mahadevan. Continuous-time hierarchical reinforcement learning. In *Proceedings of the 18th International Conference on Machine Learning*, 2001.

[8] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[9] Ronald A. Howard. Semi-markovian decision processes. In *Proceedings of International Statistical Inst.*, Ottawa, Canada, 1963.

[10] W.S. Jewell. Markov renewal programming: I. formulations, finite return models ii. infinite return models example. *Operations Research*, 11:938–971, 1963.

[11] Uffe Kjaerulff. A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence*, pages 121–129, 1992.

[12] T.-Y. Leong. *An integrated approach to dynamic decision making under uncertainty*. PhD thesis, Massachusetts Institute of Technology, 1994.

[13] Sridhar Mahadevan, Nicholas Marchalleck, Tapas K. Das, and Abhijit Gosavi. Self-improving factory simulation using continuous-time average-reward reinforcement learning. In *Proc. 14th International Conference on Machine Learning*, pages 202–210. Morgan Kaufmann, 1997.

[14] David McAllester and Luis Ortiz. Concentration inequalities for the missing mass and for histogram rule error. In *Neural Information Processing Systems 2002*, 2002.

[15] Colin McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Springer, 1998.

[16] Uri Nodelman, Christian Shelton, and Daphne Koller. Continuous time bayesian networks. In *Proceedings of the Eighteen Conference on Uncertainty in Artificial Intelligence*, pages 378–387, 2002.

[17] Martin L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley, New York, 1994.