Perfect Simulation for Markov Chains Arising From Discrete-Event Simulation

Shane G. Henderson
Dept. of Industrial and Operations Engineering
University of Michigan
1205 Beal Avenue
Ann Arbor, MI 48109-2117
shaneioe@umich.edu

Richard L. Tweedie
Division of Biostatistics, SPH
A460 Mayo Building, Mail Code 303
420 Delaware Street SE
Minneapolis, MN 55455-0378
tweedie@biostat.umn.edu

Abstract

Virtually any discrete-event simulation can be rigorously defined as a Markov chain evolving on a general state space, and under appropriate conditions, the chain has a unique stationary probability distribution. Many steady-state performance measures can be expressed in terms of the stationary probability distribution of the chain.

We would like to apply "coupling from the past" algorithms to obtain samples from the stationary probability distribution of such chains. Unfortunately, the structural properties of the chains arising from discrete-event simulations preclude the immediate application of current coupling from the past algorithms. We describe why this is the case, and extend a class of coupling from the past algorithms so that they may be applied in this setting.

1 Introduction

Virtually any discrete-event simulation can be formally described by a discrete-time Markov chain evolving on a general state space. Intuitively speaking, the chain records a snapshot of the simulation at the time at which events occur. Enough information is included in the state space to ensure that the resulting process is Markov. We will call such chains DES (discrete-event simulation) chains.

The problem of steady-state discrete-event simulation may be loosely defined as that of determining information related to the "long-run" behaviour of the system under study. There is an intimate connection between this problem and the steady-state probability distribution (if it exists) of the discrete-time Markov chain.

One cannot compute the steady-state distribution of a DES chain explicitly except in very special cases. Simulation is a natural alternative. So we are naturally led to the problem of generating samples from the stationary distribution of a DES chain. But how can one obtain such samples if the stationary distribution cannot be computed?

Recently, two methods for generating samples from the stationary distribution of a discrete-time Markov chain have emerged. Fill's algorithm [2] is, at its core, a form of rejection sampler. An alternative to Fill's algorithm is the class of methods collectively known as "coupling from the past" algorithms that were first introduced in [10, 11]. In

this paper we will explore the possibility of applying coupling from the past methods to obtain samples from the stationary distribution of a DES chain.

All of these methods operate under the same basic principle which can be described intuitively as follows. Suppose that a stationary version of the Markov chain has been running since time $-\infty$. Then at time 0, its distribution is the stationary distribution. If we can determine the state of this chain at time 0 in finite time, then that state is a sample from the stationary distribution. But how can this state be determined in finite time?

The idea is to generate sample paths of the chain from all possible starting states over the interval $[-T, 0]$, for some time $-T < 0$. If the values of all of these sample paths coincide at time 0 (coalesce), then no matter what state the chain started out in at time $-T$, it will be in the common state at time 0. Therefore, this is a sample from the stationary distribution. If the sample paths have not coalesced by time 0, then one steps back further in time, and generates sample paths from all possible starting states up to time 0, taking care to reuse the previously generated partial sample paths. This process is repeated until all paths have coalesced by time 0.

This form of coupling from the past is known as a vertical coupling. It can be shown [3] that a successful vertical coupling exists if and only if the Markov chain is uniformly ergodic. The algorithm we describe is therefore limited to this context. Other methods have been derived for dealing with chains that are not uniformly ergodic; see [13] for a survey and references.

Coupling from the past can be implemented in many ways, with the choice of implementation depending on the problem context. The method that we explore has strong connections to the theory of regenerative processes as applied in the discrete-event setting [4, 5]. The method is based on so-called minorizations of the transition kernel of the chain, and is an extension of the partitioned multigamma coupler introduced in [9].

Due to space limitations, in this paper we must content ourselves with an overview of the main ideas. Complete details may be found in [6].

This paper is organized as follows. In Section 2 we review the multigamma and partitioned multigamma couplers developed in [9]. Then, in Section 3 we explain via a simple example why these couplers do not apply immediately to DES chains. We then give an intuitive discussion of the coupler that we have developed that we call the "multistep coupler" in Section 4. Next, in Section 5 we describe some simple test examples to show that the multistep coupler performs as we might hope. We will also describe the results from, but not the details for, the example introduced in Section 3. We conclude the paper with a discussion of some interesting research questions, and some comments about how these ideas might prove useful in general discrete-event simulation.

2 Multigamma Coupling

Let $X = (X_n : n \geq 0)$ be a ϕ-irreducible aperiodic Markov chain on state space S, and suppose that the transition kernel P of X satisfies the condition

$$P(x, \cdot) \geq \lambda \varphi(\cdot) \tag{1}$$

for all $x \in S$, some probability distribution φ, and some $\lambda > 0$. Then X is a uniformly ergodic Markov chain (Theorem 16.2.3 of [8]), and has a unique stationary probability distribution π say. We wish to obtain samples from π. One approach to doing so uses the multigamma coupler [9].

Define the *residual kernel*

$$R(x,\cdot) = \frac{P(x,\cdot) - \lambda\varphi(\cdot)}{1-\lambda},$$

if $\lambda < 1$, and (arbitrarily) $R(x,\cdot) = \delta_x$, the distribution that concentrates its mass at x if $\lambda = 1$. Then

$$P(x,\cdot) = \lambda\varphi(\cdot) + (1-\lambda)R(x,\cdot). \tag{2}$$

The decomposition (2) suggests that one can generate X_1 conditional on $X_0 = x$ using the composition method (p. 448, [7]) by first generating a Bernoulli random variable Z_0 with $P(Z_0 = 1) = \lambda$. If $Z_0 = 1$, then X_1 is generated from φ, but otherwise it is generated from $R(x,\cdot)$.

The essence of the multigamma coupler is the observation that one may use the same Bernoulli random variable Z_n to generate X_{n+1} from X_n, irrespective of the value of X_n. If $Z_n = 1$, then X_{n+1} can be generated from φ, *independent of the value of* X_n. So if $Z_n = 1$, then all sample paths of X, irrespective of their value X_n at time n, couple at time $n+1$.

A formal statement of the algorithm is provided in Figure 1 below. Murdoch and Green [9] provide a statement of the multigamma coupling algorithm that is structurally quite different from the one presented in Figure 1. Their algorithm is based on a stochastic recursion, while this one is perhaps closer to the one that might be implemented. Some notes on the algorithm follow.

1. Set $T = \lfloor \log U / \log(1-\lambda) \rfloor + 1$, where $U \sim U(0,1)$
2. Generate X with distribution φ
3. for $n = -T+1$ to -1
4. Generate X from $R(X,\cdot)$
5. next n
6. Return X as a sample from π

Figure 1: Pseudocode for the multigamma coupler.

1. In line 1, we generate the first time $-T$ (before time 0) that $Z_{-T} = 1$. Since $(Z_n : -\infty < n < \infty)$ is an i.i.d. sequence of Bernoulli random variables, T has a geometric distribution, with $P(T = k) = (1-\lambda)^{k-1}\lambda$, for $k \geq 1$. The random variable U is uniformly distributed on $(0,1)$, and independent of all other quantities.

2. In line 2, we generate X_{-T+1}. Since $Z_{-T} = 1$, we generate X_{-T+1} from φ.

3. In lines 3 through 5, we successively generate $X_{-T+2}, \ldots, X_0$. Since $Z_n = 0$ for $-T < n \leq -1$, it follows that we must generate from the residual kernels.

Suppose that we resolve to apply this algorithm to the Markov chain $X = (X_n : n \geq 0)$ evolving on state space $S = \{1,2,3\}$ with transition matrix

$$P = \frac{1}{4}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 2 & 0 \end{bmatrix}.$$

This chain is clearly irreducible and aperiodic, and so has a stationary probability distribution π. Furthermore, the stationary distribution π is easily calculated, so that coupling from the past is not needed. Nevertheless, it is instructive to consider such an approach.

Suppose that we select φ to be the uniform distribution on S. Unfortunately, this chain does not satisfy the condition (1) with this choice of φ. The problem is that there is a 0 in the third row of the transition matrix P, so that we would have to take $\lambda = 0$. However, from state 3, the 2-step transition probabilities to states $1, 2, 3$ are $3/8, 3/8$ and $2/8$ respectively. So while the condition (1) is not satisfied for this chain when φ is the uniform distribution, a relaxed version of the condition is satisfied. In particular, the chain satisfies the following *minorization* condition.

Suppose that one can partition the state space S into a finite collection of sets $A_1, \ldots, A_d$, such that for each $i = 1, \ldots, d$, we have that

$$P^{m(i)}(x, \cdot) \geq \lambda_i \varphi_i(\cdot) \text{ for all } x \in A_i, \tag{3}$$

where $m(i) \geq 1$, $\lambda_i > 0$ and $\varphi_i(\cdot)$ is a probability measure on S. We call equations of the form (3) $m(i)$-minorizations. Note that (1) is a 1-minorization.

Our particular example satisfies the minorization condition by taking $d = 2$, $A_1 = \{1, 2\}$, $A_2 = \{3\}$, $m(1) = 1$, $m(2) = 2$, $\lambda_1 = \lambda_2 = 3/4$ and φ_i to be the uniform distribution for $i = 1, 2$. Many other choices are possible. For example, if we had taken φ to be uniform on the set $\{1, 2\}$ then the stronger condition (1) would have been satisfied.

The minorization assumption, together with irreducibility and aperiodicity assumptions, is sufficient to ensure that X is a Doeblin chain and hence uniformly ergodic, as verified in [6]. Such chains satisfy the condition (1) with the one step transition kernel P replaced by the m-step transition kernel P^m for some $m \geq 1$. The stationary distribution of P^m coincides with that of P, and so one could, in principle, apply the algorithm in Figure 1 to the m-step skeleton chain to obtain samples from π. There are two main reasons why this may not be recommended.

First, the m mentioned earlier may be very large, and so it may be extremely difficult to obtain expressions for the m-step transition kernel P^m. Such expressions seem to be important in implementing the multigamma coupler. Second, it may be faster [9] to develop a coupling algorithm based directly on (3) than on an m-step version of (1). This point is also discussed in [3].

So it is natural to attempt to develop a coupling from the past algorithm based on (3). The "partitioned multigamma coupler" [9] is based on a special case of (3) with $m(i) = 1$ for all i, and so we would like to apply this method in our context.

3 The Need for Multistep Coupling

Unfortunately, it has recently been shown [5] that DES chains will typically not possess a finite number of 1-minorizations that partition the state space, so that the partitioned multigamma coupler cannot be directly applied to most DES chains.

However, in [5] it is also shown that under very reasonable conditions, the chain will typically possess an m-minorization, with $m > 1$. Rather than go into the details of this general result, it is perhaps more helpful to study an example which reveals the main issues. This is the approach that we adopt in the remainder of this section. Our primary goal is to motivate the ideas to follow, by showing that we cannot expect a DES chain to possess the required finite number of 1-minorizations that partition the state space.

Consider a single-server queue with maximum capacity $Q < \infty$ where $Q \geq 2$. Customers arrive to the system according to a renewal process with interarrival time distribution function F_1. If an arriving customer finds Q customers in the system, then the customer does not enter the system. Service times are i.i.d. and independent of the arrival process, with service time distribution function F_0.

Let $X(t)$ denote the number of customers present in the system at time t, and let $T_0(t)$ ($T_1(t)$) denote the time remaining until the next service (arrival, whether the arriving customer will encounter a full system or not). If $X(t) = 0$, we arbitrarily set $T_0(t) = 0$. Let τ_n denote the time of the nth event (arrival or service) for $n \geq 1$, with $\tau_0 = 0$. Let $X = (X_n : 0 \leq n < \infty)$ denote a Markov chain with $X_n = (X(\tau_n), T_0(\tau_n), T_1(\tau_n))$, the state of the system immediately after the nth event.

Recall that we wish to develop minorizations of the transition kernel associated with the chain X.

First consider the set of states $A_1 \stackrel{\triangle}{=} \{x = (q, t_0, t_1) : q = 0, t_0 = 0, t_1 > 0\}$ corresponding to those states where the system is currently empty. The next event must necessarily be an arrival to the system, at which time both the interarrival and service time clocks are reset. So for any $x \in A_1$ and $y = (p, s_0, s_1)$, we have that

$$P(x, dy) = I(p = 1)F_0(ds_0)F_1(ds_1). \tag{4}$$

Thus, the chain X has a 1-minorization on A_1, with $\lambda_1 = 1$, and $\varphi(dy)$ given by the right-hand side of (4).

Now suppose that $x \in A = \{(q, t_0, t_1) : q = 2, 0 < t_0 < t_1\}$, the set of states where there are 2 customers in the system and the next event will be a service completion. If $x \in A$ and $y = (p, s_0, s_1)$, then

$$P(x, dy) = I(p = q - 1)F_0(ds_0)I(s_1 = t_1 - t_0).$$

Notice that the new clock reading s_1 is given by $t_1 - t_0$, the difference in clock readings at the previous step. Therefore, if φ is the probability distribution associated with any 1-minorization of P starting from $x \in A$, it immediately follows that φ must be concentrated on the set $t_1 - t_0 = c$ for some $c > 0$.

Hence, to partition the state space with a finite number of minorizations, we will at a minimum require that F_0 and F_1 are lattice distributions that are concentrated on some set $\{n\delta : 1 \leq n \leq n^*\}$ for some $\delta > 0$ and $n^* < \infty$, so that the set of possible differences in clock readings is finite. If this does not hold, then we cannot hope to partition the state space with a finite number of sets on which 1-minorizations hold.

However, under certain less restrictive conditions on the distributions F_0 and F_1 it *is* possible to partition the state space into a finite number of sets $A_1, \ldots, A_d$ together with m-minorizations where $m > 1$. Indeed, a set of such minorizations is given in [6]. The development there relies on the fact that after 2 transitions, the minorization distributions φ_i need not be concentrated on "knife-edge" sets where the difference in clock readings is fixed.

4 An Intuitive Description of the Multistep Coupler

Suppose that the condition (3) holds for a Markov chain $X = (X_n : n \geq 0)$.

Define the remainder kernels for $i = 1, \ldots, d$ by

$$R_i(x, \cdot) = \frac{P^{m(i)}(x, \cdot) - \lambda_i \varphi_i(\cdot)}{1 - \lambda_i},$$

if $\lambda_i < 1$, and (arbitrarily) $R_i(x, \cdot) = \delta_x$, the distribution that concentrates its mass at x otherwise. Then for $x \in A_i$,

$$P^{m(i)}(x, \cdot) = \lambda_i \varphi_i(\cdot) + (1 - \lambda_i) R_i(x, \cdot).$$

Let

$$m = \max_{i=1}^{d} m(i).$$

The multistep coupler is based on the following approach to generating sample paths of X. Suppose that $X_n \in A_i$. We generate a Bernoulli random variable $Z_n(i)$ with $P(Z_n(i) = 1) = \lambda_i$. If $Z_n(i) = 1$, then $X_{n+m(i)}$ is generated according to φ_i, but otherwise is generated according to $R_i(X_n, \cdot)$. We then use the 1-step transition kernel to generate $X_{n+m(i)+1}, \ldots, X_{n+m}$ (if $m(i) < m$). At time $n + m$, this process is repeated, and so the sample path is generated in steps of size m. Note that we do not generate the "intermediate" values $X_{n+1}, \ldots, X_{n+m(i)-1}$. These values are not needed for the multistep coupler.

We can generate $Z_n(i)$ for $i = 1, \ldots, d$ from a single uniform random variable U_n on $(0, 1)$ by setting $Z_n(i) = I(U_n \leq \lambda_i)$ for each i. To ensure independence of the $Z_n(i)$s across n, we require that $(U_n : -\infty < n < \infty)$ is an i.i.d. sequence.

The multistep coupler generates the first time $n = -mT_1$ before time zero that $Z_n(i) = 1$ for all $i = 1, \ldots, d$. Note that T_1 has a geometric distribution with parameter λ, where $\lambda = \min_i \lambda_i > 0$, so that $P(T_1 = j) = \lambda(1-\lambda)^{j-1}$ for $j \geq 1$. At time $-mT_1 + m$, we can be sure that every set A_i has coalesced, and so from time $-mT_1 + m$ to time 0 we only need to keep track of d paths, rather than (potentially) uncountably many.

The algorithm is then virtually identical to the partitioned multigamma coupler. In particular, we continue to step back in time in jumps of size mT_i, where $T_1, T_2, \ldots$ are i.i.d. After T_j is generated, we follow d sample paths from time $m \sum_{k=1}^{j} T_k$ to time $m \sum_{k=1}^{j-1} T_k$. We have previously generated sample paths from this latter time forward to time 0, and so we can simply follow those paths up to time 0. If all d paths originating at time $m \sum_{k=1}^{j} T_k$ coincide at time 0, then we terminate and report the common value at time 0 as a sample from π. If not, T_{j+1} is generated and the algorithm continues in a recursive fashion.

The full algorithm is given in [6].

5 Some Examples

In this section we provide 3 simple examples for which the stationary distribution is known and provide numerical results for the multistep coupler. These examples serve both as useful illustrations, and as test cases.

The first two examples are based on a Markov chain $X = (X_n : n \geq 0)$ evolving on a discrete state space $S = \{0, 1, \ldots, n\}$. This example was first described in [1]. The 1-step transition probabilities are defined by $P(x, x) = 1/2$, and $P(x, y) = 1/2n$ for $y \neq x$. The transition matrix P is doubly stochastic, and so the stationary distribution for this chain is discrete uniform on S.

This chain possesses a 1-minorization on S, where φ is the discrete uniform distribution on S, and $\lambda = (n + 1)/2n$. However, to demonstrate the features of the multistep algorithm we will consider more complicated minorizations. In both examples we take $n = 9$.

Example 1 Let $A_1 = \{0, 1, 2\}$, $A_2 = \{3, 4, 5\}$, and $A_3 = \{6, 7, 8, 9\}$.

We could take φ_i as the discrete uniform distribution on S, but it is more instructive to consider a more complicated example. Let φ_i denote the discrete uniform distribution on A_i, for $i = 1, 2, 3$. We can then take $\lambda_1 = \lambda_2 = 1/6$, and $\lambda_3 = 2/9$, while still ensuring that for $x \in A_i$, $P(x, y) \geq \lambda_i \varphi_i(\{y\})$. The residual kernels are then defined appropriately. Note here that we are taking $m(i) = 1$ for all i, so that the full multistep coupling algorithm is not explicitly required, and one could use the partitioned multigamma coupler.

We obtained 5000 samples from the stationary distribution using the multistep coupler. A 95% confidence interval for the mean number of transitions required to obtain a single sample from the stationary distribution was 31.6 ± 0.6. A chi-squared test of goodness of fit, based on 10 bins corresponding to each of the 10 states yielded a test statistic of 9.4. This is well within both the 5% and 95% percentiles (which are 3.325 and 16.919 respectively) of the chi-squared distribution with 9 degrees of freedom. Thus, it would appear that the multistep coupler is performing as it should for this problem.

Example 2 Our second example examines the same Markov chain as did Example 1, and uses the same partition sets A_1, A_2, and A_3. We use the same minorizations as in Example 1 for $x \in A_1, A_2$, but for $x \in A_3$, we take $m(3) = 2$, φ_3 to be the discrete uniform distribution on S, and $\lambda_3 = 130/162$. Indeed, it can be easily verified that $P^2(x, x) = 45/162$, and $P^2(x, y) = 13/162$ for $y \neq x$, and so it follows that for $x \in A_3$ and all $y \in S$, $P^2(x, y) \geq \lambda_3 \varphi(\{y\})$. For $x \in A_3$, it follows that $R_3(x, \cdot)$ is a point mass at x. For this example then, $m = 2$, and $\lambda = 1/6$.

We again obtained 5000 samples from the stationary distribution using the multistep coupler. A 95% confidence interval for the mean number of transitions required to obtain a single sample from the stationary distribution was 39.6 ± 0.8. A chi-squared test of goodness of fit, based on 10 bins corresponding to each of the 10 states yielded a test statistic of 5.4. This is well within both the 5% and 95% percentiles of the chi-squared distribution with 9 degrees of freedom. Thus, it would appear that the multistep coupler is again performing as it should for this problem.

Example 3 Our third example is a Markov chain $X = (X_n : 0 \leq n < \infty)$ on state space S, the half-open unit interval $[0, 1)$. The chain is defined through the recurrence

$$X_{n+1} = [X_n + U_n] \bmod 1,$$

where $(U_n : 0 \leq n < \infty)$ is an i.i.d. sequence of uniformly distributed random variables on $(-1/8, 1/8)$. For this chain, we have that for $x \in S$,

$$P(x, dy) = \begin{cases} 4\,dy & \text{for } y \in B(x, 1/8) \\ 0 & \text{otherwise,} \end{cases}$$

where $B(x, 1/8)$ is the set $\{y : y = (x + w) \bmod 1, -1/8 < w < 1/8\}$, the set of points that can be reached from $x \in S$ in one transition. It is straightforward to verify that the stationary distribution π for this chain is uniform on S.

We take $A_1 = [0, 1/8), A_2 = [1/8, 1/4), A_3 = [1/4, 1/2), A_4 = [1/2, 3/4), A_5 = [3/4, 7/8)$ and $A_6 = [7/8, 1)$.

For $x \in A_1$ and $0 \leq y < 1/8$, we have that

$$P(x, dy) = 4\,dy \geq \lambda_1 \varphi_1(dy),$$

where φ_1 is the uniform distribution on $[0, 1/8)$, and $\lambda_1 = 1/2$. Similarly, we may take φ_2 to be uniform on $[1/8, 1/4)$, φ_5 to be uniform on $[3/4, 7/8)$, and φ_6 to be the uniform distribution on $[7/8, 1)$ together with $\lambda_2 = \lambda_5 = \lambda_6 = 1/2$.

For $x \in A_3 \cup A_4$, we have that $P^2(x, \cdot)$ is a triangular distribution with parameters $(x - 1/4, x, x + 1/4)$, i.e., that

$$P^2(x, dy) = (4 - 16|y - x|)I(|y - x| < 1/4)\, dy.$$

It is then straightforward to show that with $m(3) = 2$, we can take φ_3 to be a triangular distribution with parameters $(1/4, 3/8, 1/2)$ and $\lambda_3 = 1/4$. Similarly, with $m(4) = 2$, we may take φ_4 to be a triangular distribution with parameters $(1/2, 5/8, 3/4)$ and $\lambda_4 = 1/4$.

It is easy to generate iterates from the 1-step transition kernel of this chain, as it is to generate from φ_i for $i = 1, \ldots, 6$, and from the residual kernels $R_i(x, \cdot)$ for $i = 1, 2, 5, 6$. However, generating from the residual kernels $R_3(x, \cdot)$ and $R_4(x, \cdot)$ is somewhat nontrivial. We therefore employ an acceptance rejection method for generating from these distributions, using $P^2(x, \cdot)$ as a proposal distribution.

We again obtained 5000 samples from the stationary distribution using the multistep coupler. A 95% confidence interval for the mean number of transitions required to obtain a single sample from the stationary distribution was 101 ± 2. A chi-squared test of goodness of fit, based on 10 bins corresponding to the 10 subintervals $[k/10, (k + 1)/10)$, $k = 0, 1, \ldots, 9$ yielded a test statistic of 7.7. This is well within both the 5% and 95% percentiles of the chi-squared distribution with 9 degrees of freedom. Thus, it would appear that the multistep coupler is again performing as it should for this problem.

We have also conducted experiments on the DES chain described in Section 3. For full details the reader is referred to [6]. The results there can be summarized by stating that for a queue capacity of size Q, a total of $d = 4Q$ minorizations of the form (3) are required to partition the state space. For $Q = 10$, the expected number of transitions required to identify a sample from the stationary distribution is approximately 1000. Hence, obtaining perfect samples in that case is computationally expensive relative to simulating a single transition of the chain.

6 Conclusions and Future Research

The detailed results in [6] suggest that it is certainly possible to obtain a sample from the stationary distribution of the embedded Markov chain in certain cases. The principal difficulty in obtaining such results lies in constructing the required minorizations. This appears to be a highly nontrivial task, and is, at least currently, a barrier to widespread application. Nevertheless, we have demonstrated that in principle, it is possible to obtain samples from the stationary distribution of the embedded chain.

This naturally leads to a strongly related question, namely, can one obtain samples from the stationary distribution (if it exists) of the continuous-time Markov chain that, sampled at event times, gives rise to the discrete-time Markov chain we have analyzed? If so, we would have a method for solving the initial transient problem, at least for a class of models. Such questions are related to Palm theory for point processes; see, for example, [12].

It is also interesting to ask whether one can avoid the uniform ergodicity requirement implicit in the use of vertical couplings. Some general techniques have been developed to

combat this difficulty (an overview and references are given in [13]), but whether these techniques can be successfully applied to this setting remains to be seen.

As noted earlier, there is a strong connection between the theory presented here, and the theory of regenerative simulation of discrete-event dynamic systems (see [5]). Both approaches rely, in the general case, on developing minorizations for m-step transition kernels of a DES chain. If one can apply the multistep coupler developed here, then it appears that one can apply regenerative steady-state simulation methodology. The converse is not necessarily true, because the requirements of the regenerative method are weaker than those of the multistep coupler, at least as presented here.

Given the very attractive statistical properties of estimators constructed from regenerative arguments, one might very well ask why we would want perfect samples from the stationary distribution of the embedded chain. This is a reasonable objection. However, if we are able to leverage these results to obtain perfect samples from the stationary distribution $\tilde{\pi}$ say, of the continuous-time Markov chain alluded to above, then there is certainly potential gain.

To see why, suppose that we are trying to estimate $\alpha = \int f(x)\tilde{\pi}(dx)$, where $\tilde{\pi}$ is the stationary distribution of the continuous-time Markov chain $X = (X(t) : t \geq 0)$, and f is a cost function defined on the state space of X. A regenerative estimator of this quantity is based on a single run of the chain X, and takes the form of a ratio. Typically then, it suffers from "ratio estimation" bias. If this ratio estimation bias is severe and cannot be sufficiently mitigated with bias-reduction techniques, then an alternative approach, based on coupling from the past, might be preferred.

In such a method, we would generate a moderate number, n say, of independent samples from $\tilde{\pi}$ using coupling from the past methods, and then use these samples as initial points in n fixed runlength simulations of the continuous-time Markov chain X. If $Y_1, \ldots, Y_n$ are n sample path averages defined via $Y_i = t^{-1} \int_0^t f(X^i(s))\, ds$, where X^i is the ith replication of the sample path of the chain X and t is the simulation runlength, then the Y_is are i.i.d. *unbiased* estimates of α. We can then estimate α by the average of the Y_is, and confidence intervals are easily constructed. The unbiased nature of this estimator, and the ease with which we can obtain confidence intervals, comes at the cost of generating the n initial points using coupling from the past.

These properties of a coupling from the past type estimator are certainly attractive, and suggest that there is potential value in this approach.

References

[1] Andradóttir S., Calvin, J. M., and Glynn, P. W. (1995) Accelerated regeneration for Markov chain simulations. *Probab. Engrg. Inform. Sci.* **9,** 497–523.

[2] Fill, J. A. (1998). An interruptible algorithm for perfect sampling via Markov chains. *Annals of Applied Probability* **8** 131–162.

[3] Foss, S. G., and Tweedie, R. L. (1998). Perfect simulation and backward coupling. *Communications in Statistics Stochastic Models.* **14** 187–203.

[4] Henderson, S. G., and P. W. Glynn (1999). Derandomizing variance estimators. *Operations Research* **47** 907–916.

[5] Henderson, S. G., and P. W. Glynn (2000). Regenerative steady-state simulation of discrete-event systems. *Under revision.*

[6] Henderson, S. G. and R. L. Tweedie. (2000). Discrete-event steady-state simulation and coupling from the past. *In Preparation.*

[7] Law, A. M., and Kelton, W. D. (2000) *Simulation Modeling and Analysis*. 3rd edn. McGraw Hill, Boston.

[8] Meyn, S. P., and Tweedie, R. L. (1993) *Markov Chains and Stochastic Stability.* Springer-Verlag, London.

[9] Murdoch, D. J., and Green, P. J. (1998) Exact sampling from a continuous state space. *Scand. J. Statist.* **25,** 483–502.

[10] Propp, J. G., and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9** 223–252.

[11] Propp, J. G., and Wilson, D. B. (1998). How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *Journal of Algorithms* **27** 170–217.

[12] Thorisson, H. (2000). *Coupling, Stationarity and Regeneration.* Springer, New York.

[13] Wilson, D. B. (2000). Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP). In N. Madras, ed., *Monte Carlo Methods, volume 26 of Fields Institute Communications*, p. 141–176. American Mathematical Society.