Probability in the Engineering and Informational Sciences, **13**, 1999, 147–167. Printed in the U.S.A.

TRANSIENT SIMULATION VIA EMPIRICALLY BASED COUPLING

EUGENE W. WONG, PETER W. GLYNN, AND DONALD L. IGLEHART*

Department of Engineering-Economic Systems and Operations Research
Stanford University
Stanford, California 94305-4023

In this paper we consider the use of coupling ideas in efficiently computing a certain class of transient performance measures. Specifically, we consider the setting in which the stationary distribution is unknown, and for which no exact means of generating stationary versions of the process is known. In this context, we can approximate the stationary distribution from empirical data obtained from a first-stage steady-state simulation. This empirical approximation is then used in place of the stationary distribution in implementing our coupling-based estimator. In addition to the empirically based coupling estimator itself, we also develop an associated confidence interval procedure.

1. INTRODUCTION

Let $X = \{X(t) : t \geq 0\}$ be a stochastic process that represents the output of a simulation. This paper is concerned with the efficient computation, via simulation, of transient performance measures of the form

$$\alpha = E\beta,$$

where β is defined by

$$\beta = \int_{[0,\infty)} f(X(t))G(dt). \tag{1.1}$$

We assume, in Eq. (1.1), that f is a given real-valued function defined on the state space S of X, and G is a given finite measure.

*This research was supported by the Army Research Office under Contract no. DAAG55-97-1-0377.

 0269-9648/99 $12.50

A number of important transient performance measures can be represented in the form of Eq. (1.1).

Example 1: To compute $\alpha = Ef(X(t))$, set $G(ds) = \delta_t(ds)$ for $s \geq 0$, where $\delta_t(\cdot)$ is a unit mass at the point t.

Example 2: To compute the expected cumulative "cost" of running the system to time t, we let $G(ds) = I(0 \leq s \leq t)\, ds$. (We interpret $f(x)$ as the rate at which cost accrues when X occupies state x.)

Example 3: Assume, as in Example 2, that $f(x)$ represents the rate at which the system accrues cost when X is in state x. If

$$\alpha = E \int_{[0,\infty)} e^{-\gamma t} f(X(t))\, dt,$$

the infinite-horizon γ-discounted cost, this performance measure may be represented as a special case of Eq. (1.1) by setting $G(dt) = e^{-\gamma t}\, dt$ for $t \geq 0$.

In a previous paper (Glynn and Wong [7]) we showed how coupling ideas can be used to efficiently compute performance measures of the form in Eq. (1.1) (for general background on coupling, see Lindvall [8]). Specifically, suppose that X is a strong Markov process that possesses a unique stationary distribution π. If we are able to generate variates from π, this gives us the ability to simulate a stationary version X^* of the process (by initiating the Markov process X with its stationary distribution π). Suppose that we simultaneously simulate X in such a way that X and X^* "couple." In other words, X and X^* are jointly simulated so that the random time $T = \inf\{t \geq 0 : X(t) = X^*(t)\}$, known as the "coupling time," is almost surely finite. Note that for $t \geq T$, we may set $X(t) = X^*(t)$ without changing the distribution of X. So

$$\begin{aligned}\alpha &= E \int_{[0,T)} f(X(t))G(dt) + E \int_{[T,\infty)} f(X^*(t))G(dt) \\ &= E \int_{[0,T)} [\, f(X(t)) - f(X^*(t))]G(dt) + E \int_{[0,\infty)} f(X^*(t))G(dt).\end{aligned}$$

Note that $Ef(X^*(t)) = Ef(X^*(0))$ for $t \geq 0$ by stationarity of X^*. Consequently, if $Ef(X^*(0))$ can be computed either analytically or numerically, the above argument ensures that $\alpha = E\Gamma$, where

$$\Gamma = \int_{[0,T)} [\, f(X(t)) - f(X^*(t))]G(dt) + gEf(X^*(0)). \qquad \textbf{(1.2)}$$

The quantity g appearing in Eq. (1.2) is the total mass of G. This identity proves that α can be computed by generating i.i.d. replicates of the random variable Γ.

Glynn and Wong [7] study the efficiency of the above coupling-based simulation algorithm in the context of Examples 1 through 3. For Examples 1 and 2, it is

shown that as $t \to \infty$, the coupling-based estimator dominates the naive estimator algorithm (based on replicating β). The coupling-based estimator is also shown to improve efficiency in the setting of Example 3. In addition, various extensions of the methodology are provided, including an extension to the setting in which $Ef(X^*(0))$ must be computed by simulation.

All the methodologies described in [7] presume the ability to generate variates from the stationary distribution π. While there are a number of important applications in which this is possible (and for which the transient behavior is intractable), most real-world simulations do not possess this property. This paper is devoted to extending the above coupling-based methodology to this latter setting. The idea is first to simulate X over some long time horizon and then compute an empirical approximation to π from the simulation data. By using the empirical approximation in place of π, we can then simulate an approximately stationary version of X. Furthermore, $Ef(X^*(0))$ can be estimated via the time-average obtained from the initial simulation. If a suitable coupling exists, we can then generate an approximation to the random variable Γ; see Eq. (1.2). By generating replications of this approximation to Γ, we can then compute an estimator for α. We refer to this estimator as our empirically based coupling estimator for α.

The main contributions of this paper include:

1. development of a central limit theorem (CLT) for the empirically based coupling estimator for α (Theorem 1);
2. development of a confidence interval procedure for our empirically based coupling estimator (Theorem 2);
3. discussion of conditions under which the empirically based coupling estimator is to be preferred to the conventional estimator (based on averaging i.i.d. replications of β);
4. introduction of several different empirical approximations to the distribution π that are suitable for use in our empirically based coupling estimator.

The empirically based coupling estimator is carefully described in Section 2, and its basic properties discussed; proofs are deferred to Section 4. Our computational results are discussed in Section 3.

2. DESCRIPTION OF THE ALGORITHM AND MAIN RESULTS

To describe the above algorithm, we start with a given computer budget c. A proportion r $(0 < r < 1)$ is allocated to generating an empirical approximation to π, whereas $(1 - r)$ is allocated to generating the replicates of the approximation to the random variable Γ. So $c_1 = rc$ and $c_2 = (1 - r)c$ are the amounts of computer time allocated to the first and second "stage" of the procedure. To simplify our analysis and notation, we assume that simulated time equals computer time, so that with t units of computer time, exactly t time units of X can be simulated. With (possibly) a trivial change of time scale, this assumption, while not true in an exact sense, is at least approximately true in most real-world simulations.

The first stage of the procedure requires simulating X to time c_1, at which time an empirical approximation π_c to π is obtained from the simulated data. For example, if X has discrete state space, a suitable empirical approximation is given by the empirical distribution

$$\pi_c(\cdot) = \frac{1}{c_1}\int_0^{c_1} I(X(s) \in \cdot)\, ds. \tag{2.3}$$

For continuous state space processes, we will need to use other approximations to π; we will discuss this point in further detail later. In addition, the first stage provides the estimator

$$\gamma(c_1) = g\int_S f(y)\pi_c(dy)$$

to the quantity $gEf(X^*(0))$ appearing in Eq. (1.2).

In the second stage we independently generate approximations to the random variable Γ. Let μ be the initial distribution of the process X for which we are trying to compute the unknown transient performance measure α. Then, let X_{11} and X_{12} be two versions of X, in which X_{11} is initiated with distribution μ and X_{12} is initiated with distribution π_c. Assuming that X_{11} and X_{12} are simulated in such a way that they couple at time T_1, this yields an approximation $\Gamma_1(c)$ to the random variable Γ, namely

$$\Gamma_1(c) = \int_{[0,T_1)} [\, f(X_{11}(t)) - f(X_{12}(t))]G(dt) + \gamma(c_1).$$

We then expend the remaining c_2 time units of our computer budget, independently, generating additional replicates $(X_{21}, X_{22}),(X_{31}, X_{32}),\ldots$ of (X_{11}, X_{12}) until our budget is exhausted. Let $N(c_2)$ be the number of such replicates produced in the remaining c_2 time units. Then, our empirically based coupling estimator is given by

$$\alpha(c) \stackrel{\text{def}}{=} \frac{1}{N(c_2)}\sum_{i=1}^{N(c_2)} \Gamma_i(c). \tag{2.4}$$

Something needs to be said about how to construct the coupling time T_i for (X_{i1}, X_{i2}). If the state space is discrete and X is an irreducible, positive recurrent, continuous-time Markov chain, then we may simulate X_{i1} and X_{i2} independently until they (necessarily) meet at time T_i, after which we set $X_{i1} = X_{i2}$. If the state space is continuous, X_{i1} and X_{i2} may never meet if they are simulated independently. More sophisticated couplings may be necessary in these circumstances [7].

In order to describe our main results, we need to be precise about the assumptions underlying our analysis. We shall assume that X is an S-valued Markov process, having initial distribution μ, and possessing stationary transition probabilities. We require that S be a complete separable metric space. Through suitable use of "supplementary variables," virtually any discrete-event simulation may be viewed as such a process; see Glynn [5] for details. (Note that Euclidean space $\Re^d$ and

finite/countably infinite state spaces are complete separable metric spaces.) We assume that X has right continuous paths with left limits, and possesses a stationary distribution π. Thus, $X = \{X(t) : t \geq 0\} \in D$, the space of right-continuous S-valued functions with left limits.

Our most critical assumption is that for any initial state $x \in S$, the process initiated at x can be coupled to the process initiated with distribution μ. A more careful statement of this assumption involves letting $X_i : D \times D \to D$ be the coordinate projections defined by $X_i(x_1, x_2) = x_i$, for $i = 1,2$. Let $T = T(x_1, x_2) = \inf\{t \geq 0 : x_1(t) = x_2(t)\}$. Our coupling assumption demands the existence of a family $(P_x : x \in S)$ of probability distributions on $D \times D$ such that

i. $P_y(B)$ is measurable in y for each (measurable) B
ii. $P_x(X_1 \in \cdot) = P(X \in \cdot)$, $x \in S$
iii. $P_x(X_2 \in \cdot) = P(X \in \cdot | X(0) = x)$, $x \in S$
iv. $P_x(T < \infty) = 1$, $x \in S$.

Assumptions ii, iii, and iv assert that X_1 has distribution of X associated with initial distribution μ, X_2 has the distribution of X initiated at x, and the coupling time is finite. (In ii above, it should be noted that the distribution P always denotes a distribution under which X has initial distribution μ.)

For any two probabilities P_1 and P_2 defined on the same sample space, let

$$\|P_1 - P_2\| \stackrel{\text{def}}{=} \sup_A |P_1(A) - P_2(A)|$$

be the "total variation distance" between P_1 and P_2.

It turns out that the hypotheses we have stated imply that X is recurrent in a certain sense.

DEFINITION 1: *A Markov process $Z = \{Z(t) : t \geq 0\}$, taking values in S, is said to be a Harris recurrent Markov process if there exists a probability measure η such that whenever $\eta(A) > 0$,*

$$P\left(\int_0^\infty I(Z(t) \in A)\, dt = +\infty | Z(0) = z\right) = 1$$

for all $z \in S$.

PROPOSITION 1: *The Markov process X is Harris recurrent. Furthermore,*

$$\|P(X(t) \in \cdot | X(0) = x) - \pi(\cdot)\| \to 0 \qquad \textbf{(2.5)}$$

as $t \to \infty$, for each $x \in S$.

Note that Proposition 1 establishes that our hypotheses preclude X from having periodic behavior.

To analyze our algorithm, we now define our empirically based estimator in more rigorous terms. In addition to simulating the process X up to time c_1, the

algorithm further involves the D-valued random elements and coupling times $(X_{11}, X_{12}, T_1, X_{21}, X_{22}, T_2, \ldots)$ associated with the remaining budget c_2. We shall assume that our sample space Ω is sufficiently rich so as to support, for each $c > 0$, $(X_{11}(c), X_{12}(c), T_1(c), X_{21}(c), X_{22}(c), T_2(c), \ldots)$. The algorithm also requires the existence of an empirical approximation π_c to π. For $c \geq 0$, let $\pi_c = \{\pi_c(B) : (\text{measurable}) \ B \subset S\}$ be a set-indexed process such that:

a. π_c is measurable with respect to $\{X(s) : 0 \leq s \leq c_1\}$. (In other words, π_c is a (deterministic) function of $\{X(s) : 0 \leq s \leq c_1\}$.)

b. $\pi_c(\cdot)$ is a probability on S.

The probability P defined on Ω is chosen so that X has the same distribution as before. (In particular, X is Markov with initial distribution μ.) Furthermore, we choose P so that

c. $P((X_{i1}(c), X_{i2}(c), T_i(c)) \in A_i, 1 \leq i \leq n | X) = \prod_{i=1}^{n} \int_S \pi_c(dx) P_x((X_1, X_2, T) \in A_i)$.

So conditional on X, $\{(X_{i1}(c), X_{i2}(c), T_i(c)) : 1 \leq i \leq n\}$ are i.i.d., with $X_{i1}(c)$ having the distribution of X, $X_{i2}(c)$ having the distribution of X initiated under π_c, and $T_i(c)$ representing the corresponding coupling time. Set

$$\Gamma_i(c) = \int_{[0, T_i(c))} [f(X_{i1}(c,t)) - f(X_{i2}(c,t))] G(dt) + \gamma(c_1),$$

where $\gamma(c_1) = g \int_S \pi_c(dy) f(y)$. Let $s = \sup\{t \geq 0 : G(t) < G(\infty)\}$ be the right end-point of the support of G. Note that computing $\Gamma_i(c)$ requires simulating $X_{i1}(c)$ and $X_{i2}(c)$ up to time $T_i(c) \wedge s$. Thus $\chi_i(c) = 2(T_i(c) \wedge s)$ is the computer time involved in generating $\Gamma_i(c)$ (incremental to that associated with $\gamma(c_1)$). Hence, $N(c_2) = \max\{n \geq 0 : \chi_1(c) + \cdots + \chi_n(c) \leq c_2\}$ is the number of $\Gamma_i(c)$'s generated within the second stage's budget c_2. Our empirically based coupling estimator can then be defined via Eq. (2.2).

Our first major result is a central limit theorem (CLT) for $\alpha(c)$. In preparation for this set

$$R = \int_{[0,T)} [f(X_1(t)) - f(X_2(t))] G(dt),$$

$g_p(x) = E_x R^p$, and $h_p(x) = E_x (T \wedge s)^p$. We assume that

A1. $g_2(x) < \infty$ for $x \in S$, $h_2(x) < \infty$ for $x \in S$, and

$$\int_S |f(x)| \pi(dx) < \infty.$$

Let $k(x) = f(x) g + E_x R$. Observe that Eq. (1.2) asserts that $\int_S k(x) \pi(dx) = \alpha$. We require the existence of a finite (deterministic) constant σ_1 such that

A2. $\sqrt{c_1}(\int_S k(y)\pi_c(dy) - \alpha) \Rightarrow \sigma_1 N(0,1)$ as $c_1 \to \infty$, where $\Rightarrow$ denotes weak convergence.

As for the empirical approximation π_c, we require that π_c satisfy

A3. $\|\pi_c - \pi\| \to 0$ a.s.;

A4. $\int_S g_2(x)\pi_c(dx) \to \int_S g_2(x)\pi(dx)$ a.s.;

A5. $\int_S h_2(x)\pi_c(dx) \to \int_S h_2(x)\pi(dx)$ a.s.; as $c \to \infty$.

We will discuss conditions A3–A5 in greater detail later. For A2, we may invoke the CLT for Harris processes, by virtue of Proposition 1; see Glynn [6] for such a CLT. Let

$$P^*(\cdot) = \int_S \pi(dx)P_x(\cdot).$$

THEOREM 1: *Assume* A1–A5. *Then*, $\lambda^{-1} = 2E^*(T \wedge s) < \infty$ *and* $\sigma_2^2 = \text{var}^* R < \infty$, *where* $E^*(\cdot)$ *and* $\text{var}^*(\cdot)$ *are the expectation and variance operators associated with* P^*, *respectively*. *Furthermore*,

$$\sqrt{c}(\alpha(c) - \alpha) \Rightarrow \sigma N(0,1)$$

as $c \to \infty$, *where*

$$\sigma^2 = \frac{\sigma_1^2}{r} + \frac{\lambda^{-1}\sigma_2^2}{1-r}. \tag{2.6}$$

The quantity $\lambda^{-1}\sigma_2^2$ is precisely the asymptotic variance constant that appears in the CLT when exact coupling is applicable. Specifically, this corresponds to the case when $Ef(X^*(0))$ can be computed exactly, and it is possible to generate variates from π. It should also be noted that the variance constant σ^2 appearing in Theorem 1 is not the same as that which appears when $Ef(X^*(0))$ is unknown, with the possibility of exact generation of variates from π. In this latter case, $Ef(X^*(0))$ must be estimated from a first-stage steady-state simulation of X. However, the first-stage plays no role in initializing the second-stage coupling-based replications. As a consequence, it turns out that the asymptotic variance in this latter setting has the same form as Eq. (2.6), with the sole modification being that σ_1^2 is then the time-average variance constant for the function $gf(x)$ (instead of $k(x)$).

We now perform an asymptotic analysis, in order to get a sense of when our empirically based coupling estimator is preferable to conventional simulation (in which i.i.d. replicates of β are averaged). We consider first Example 1, in which $\beta = f(X(t))$. In this case, we typically expect $E_x R \to 0$ as $t \to \infty$; this is guaranteed if f is bounded. Also, $\sigma_2^2 \to 0$ as $t \to \infty$ is to be expected. Hence, for large t, σ^2 ought to be roughly equal to the time-average variance constant of $f(X(t))$, divided by r. On the other hand, the corresponding variance constant for the conventional estimator grows linearly in t as $t \to \infty$ [7]. Hence, the empirically based coupling estimator is a clear winner for t large.

As for Example 2, in which β is a cumulative cost over $[0,t]$, both the empirically based coupling estimator and conventional estimator have respective variance constants that typically grow in proportion to t^2. Here, no clear winner exists, even for large t. Finally, in the context of Example 3, β cannot be generated in finite time, so the conventional estimator is infeasible. (However, alternatives to the empirically based coupling estimator exist; see Fox and Glynn [4].) Thus, the empirically based coupling estimator is again a winner.

We now turn to the question of producing confidence intervals for α, based on the estimator $\alpha(c)$ that we have introduced in this paper. The challenge here is dealing with σ_1^2, that involves the unknown function k. To deal with this problem, we use the method of batch means. To validate this procedure, we require stronger hypothesis on the empirical approximation π_c to π.

A6. For each $c > 0$, there exists a family of probabilities $\{\nu(c,x,\cdot) : x \in S\}$ such that

$$\pi_c(B) = \frac{1}{c_1}\int_0^{c_1} \nu(c, X(s), B)\, ds.$$

A7. There exists a finite (deterministic) constant σ_1 and a standard Brownian motion B such that

$$\sqrt{c}\left(\frac{1}{c}\int_0^t \int_S k(y)\nu(c, X(s), dy)\, ds - \alpha t\right) \Rightarrow \sigma_1 B(t)$$

as $c \to \infty$. Here $\Rightarrow$ denotes weak convergence in the topology associated with the space D_R of real-valued right-continuous functions with left limits.

Assumption A7 is, from a practical standpoint, only slightly stronger than A2. Observe that we can generate $X_{12}(c,0), X_{22}(c,0), \ldots$ from the distributions

$$\nu(c, X(cU_1), \cdot), \nu(c, X(cU_2), \cdot), \ldots,$$

where $U_1, U_2, \ldots$ is an independent sequence of i.i.d. uniform r.v.'s on $[0,1)$. Then, for $1 \le i \le m$, let

$$\alpha_i(c) = \frac{m}{N(c_1)} \sum_{j=1}^{N(c_1)} R_j(c) I\left(\frac{i-1}{m} \le U_j < \frac{i}{m}\right) + \frac{gm}{c_1}\int_{(i-1)c_1/m}^{ic_1/m} f(X(s))\, ds.$$

Note that

$$\alpha(c) = \frac{1}{m}\sum_{i=1}^{m} \alpha_i(c).$$

Let

$$s(c) = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(\alpha_i(c) - \alpha(c))^2}.$$

THEOREM 2: *Fix $m \geq 2$. Under Assumptions* A1 *and* A3 *to* A7, *it follows that*

$$\sqrt{m}\left(\frac{\alpha(c) - \alpha}{s(c)}\right) \Rightarrow t_{m-1}$$

as $c \to \infty$, where t_{m-1} is a Student-t r.v. with $m-1$ degrees of freedom.

Theorem 2 implies that if z is a solution to $P(-z \leq t_{m-1} \leq z) = 1 - \delta$, then

$$\left[\alpha(c) - \frac{zs(c)}{\sqrt{m}}, \alpha(c) + \frac{zs(c)}{\sqrt{m}}\right]$$

is an approximate $100(1-\delta)\%$ confidence interval for α when c is large. Hence, Theorem 2 yields a confidence interval methodology for our empirically based coupling estimator.

A natural question that arises here is the computation of the optimal value of r. An easy calculation shows that the value of r that minimizes σ^2 is

$$r^* = \left(1 + \sqrt{\frac{\lambda^{-1}\sigma_2^2}{\sigma_1^2}}\right)^{-1}.$$

Because we do not know k explicitly, it is unclear how to estimate σ_1^2 consistently. Instead, we can use the following (reasonable) heuristic procedure. If we wish to compute α using an approximately optimal value of r, we first do a couple of "trial runs" using $r = \frac{1}{3}$ and $r = \frac{2}{3}$, respectively, and the same computer budget c. From the two trial runs, we obtain two confidence interval half-widths H_1 and H_2, respectively. Then,

$$H_1^2 \approx \frac{t^2}{c}\left(3\sigma_1^2 + \frac{3}{2}\lambda^{-1}\sigma_2^2\right)$$

$$H_2^2 \approx \frac{t^2}{c}\left(\frac{3}{2}\sigma_1^2 + 3\lambda^{-1}\sigma_2^2\right).$$

Hence,

$$\frac{\lambda^{-1}\sigma_2^2}{\sigma_1^2} \approx \frac{2H_2^2 - H_1^2}{2H_1^2 - H_2^2}.$$

Consequently,

$$r^* \approx \left(1 + \sqrt{\frac{2H_2^2 - H_1^2}{2H_1^2 - H_2^2}}\right)^{-1}.$$

Thus, we can now do our "production run," using computer budget c and $r = r^*$ as obtained through the above formula.

We conclude this section with a discussion of empirical approximations π_c to π that satisfy our hypotheses. We consider first the case in which S is discrete.

PROPOSITION 2: *Suppose S is either finite or countably infinite. If*

$$\pi_c(B) = \frac{1}{c_1}\int_0^{c_1} I(X(s) \in B)\, ds,$$

then $\|\pi_c - \pi\| \to 0$ a.s. as $c \to \infty$.

We turn next to the verification of A3 when S is continuous. In this setting, it is typically untrue that the empirical distribution itself will converge to π a.s. in the sense of total variation. So we use a different empirical approximation here. We assume that there exists a probability η and $0 < t_0 < \infty$ such that for each $x, y \in S$,

$$P(X(t_0) \in dy | X(0) = x) = p(x,y)\eta(dy), \tag{2.7}$$

where $p : S \times S \to \Re$ is a known (density) function. It is rarely the case that such a density will be computable in continuous time (unless the entire transient distribution is known). However, in discrete time, the situation is quite different.

Specifically, the theory developed in this paper for continuous time Harris recurrent processes extends easily to sequences $\{X_n : n \geq 0\}$ that are Harris recurrent Markov chains. In particular, by embedding $\{X_n : n \geq 0\}$ in continuous time via $X(t) = X_{\lfloor t \rfloor}$ for $t \geq 0$, the estimator $\alpha(c)$ and corresponding confidence interval methodology continue to be valid for simulations of such chains. For such a discrete-time chain, $P(X_1 \in dy | X_0 = x)$ can easily be computed in closed form. So, in discrete time, the above assumption is quite reasonable.

PROPOSITION 3: *Suppose $p : S \times S \to \Re$ is continuous and bounded. If*

$$\pi_c(B) = \frac{1}{c_1}\int_0^{c_1}\int_B p(X(s), y)\eta(dy)\, ds,$$

then $\|\pi_c - \pi\| \to 0$ a.s. as $c \to \infty$.

Note that A4–A7 can easily be verified for the empirical approximations we have proposed above. In particular, A4 and A5 require invoking the strong law for Harris processes, whereas A7 involves using the functional CLT for Harris processes.

Another class of empirical approximations, applicable when $S = \Re^d$, is based on setting

$$\nu(c, x, dy) = P(x + b(c)N(0,1) \in dy),$$

where $b(c) \to 0$ at a suitable rate as $c \to \infty$. If $\{b(c) : c > 0\}$ is chosen appropriately, the corresponding empirical approximation has a density that converges a.s. to the

stationary density; see Devroye [3] for details in the i.i.d. context. Verifying A4–A7 again comes down to invoking laws of large numbers and central limit theorems for Harris processes.

3. NUMERICAL EXPERIMENTS

In this section, we compare the performance of the empirically based coupling estimator and the conventional estimator by the means of numerical experiments. In addition, we report the performance of the coupling-based estimator for the purpose of comparison. The experiments are done on models for which both the transient means and stationary distributions are available analytically. The models we have chosen are the $M/M/\infty$ queue-length process, the $M/M/1$ waiting time process, and the $AR(1)$ autoregressive process.

In order to provide a fair comparison, we assign the same computer budget to each of the estimators. We assume the computer time can be taken as equivalent to the simulation time. This assumption simplifies matters because of the machine-dependent subtleties that arise by explicitly timing each of these algorithms.

The coupling-based estimator and its confidence interval for α were introduced in [7]. The empirically based estimator and its confidence intervals were constructed using the batch means method described in the last section. The number of batches we used in each of the experiments is set to 20. Before the actual simulation of the empirically based estimator, we perform a pilot run to determine the optimal value of r based on the heuristic method given in the last section.

Since H_1^2 and H_2^2 may not be good approximations for

$$\frac{t^2}{c}\left(3\sigma_1^2 + \frac{3}{2}\,\lambda^{-1}\sigma_2^2\right)$$

and

$$\frac{t^2}{c}\left(\frac{3}{2}\,\sigma_1^2 + 3\lambda^{-1}\sigma_2^2\right),$$

respectively, we may end up having

$$\frac{2H_2^2 - H_1^2}{2H_1^2 - H_2^2} < 0.$$

In that case, we simply set r to some default value, for example, 0.5. Note that we use the same r to generate our estimator in each replication.

We then replicate the experiment on each estimator a number of times. The mean error, the mean square error, and the coverage of confidence intervals relative to the true mean α is calculated for each estimator. The mean error is given by averaging the difference of the estimator and the true mean for all replications, the mean square error is the average of the square of the difference of the estimator and

the true mean for all replications, and the coverage is defined as the number of confidence intervals that contain α divided by the total number of confidence intervals generated. We expect 1σ, 2σ, and 3σ coverages to correspond roughly to 68%, 95%, and 99.7% of the confidence intervals covering α.

The first model we consider is the $M/M/\infty$ queue-length process. This is a birth-death process $X = \{X(t) : t \geq 0\}$ that has birth rates $\lambda_n = \lambda$ and death rates $\mu_n = n\mu$ for $n \geq 0$. In this case, X lives in a discrete state space. It is well known that if $X(0) = 0$, then $X(t)$ is Poisson distributed with parameter $(\lambda(1 - \exp(\mu t))/\mu)$. Suppose we are interested in estimating the quantity $\alpha = E(X(t))$. This corresponds to setting $f(x) = x$ and $G(dx) = \delta_t(dx)$ in Eq. (1.1). Then

$$\alpha = Ef(X(t)) = \frac{\lambda(1 - \exp(-\mu t))}{\mu}.$$

Let c be the total computer budget and $c_1 = rc$ $(r \in (0,1))$. The coupling used here is the independent coupling; we simulate the two processes independently until they couple. Tables 1 and 2 compare the absolute errors, the mean square errors, and the coverage of the conventional estimator, $\alpha_0(c)$, the coupling-based estimator, $\alpha_1(c)$, and the empirically based coupling estimator, $\alpha_2(c)$.

The performance of the conventional estimator, as measured by the mean square error, degrades as t gets large, whereas the performance of the coupling-based estimator $\alpha_1(c)$ improves as $t \to \infty$. It should come as no surprise that the variance of $\alpha_1(c)$ goes to zero as $t \to \infty$. The empirically based coupling estimator $\alpha_2(c)$ does not show significant change as $t \to \infty$. As discussed in the previous section, the variance associated with the estimator, σ^2, converges to a constant as $t \to \infty$. Since the empirically based coupling estimator does not degrade, its performance is better than the conventional estimator for large t as shown in the experiments. In terms of confidence intervals coverage, all three estimators give reasonable results.

The next model we study is the $M/M/1$ waiting time process, $W = \{W(i) : i \geq 0\}$ with traffic intensity $\rho < 1$. It is an irreducible, positive recurrent, discrete-time Markov chain living on a continuous state space, $\Re^+$. Here we are interested in estimating $EW(n)$ for $n > 0$. Again, we use independent coupling, with coupling time

$$T_i(c) = \inf\{m \geq 0 : W_{11}(c,m) = W_{12}(c,m) = 0\}.$$

Since the aperiodic chain W visits state 0 infinitely often, we have $P(T_i(c) < \infty) = 1$.

The expected value $EW(n)$ can be obtained as follows. Let p be the probability that the existing customer will be served before a new customer arrives. So, $p = \mu/(\lambda + \mu)$; the corresponding probability that a new customer will arrive before an existing customer is served is $q = 1 - p$. Let P_{ij} be the probability that, if there are currently i customers in the system, there will be j customers in the system just prior to the arrival of the next customer. Therefore, for each i,

TABLE 1. $M/M/\infty$ Queue-Length Process with $\lambda = 0.5$, $\mu = 1.0$

t	Estimator	Replications	c	r	Error	MSE	1σ	2σ	3σ
0.5	α_0	100	20000	0	−0.000075670	0.000006657	0.54	0.93	0.99
0.5	α_1	100	20000	0	−0.000178395	0.000005022	0.68	0.99	1.00
0.5	α_2	100	20000	0.44	−0.000157398	0.000024756	0.70	0.97	0.99
1.0	α_0	100	20000	0	0.000177221	0.000013352	0.76	0.99	0.99
1.0	α_1	100	20000	0	−0.000006267	0.000006702	0.68	0.95	0.99
1.0	α_2	100	20000	0.40	−0.00015954	0.00007541	0.68	0.92	0.99
2.0	α_0	100	20000	0	−0.001160358	0.000035513	0.76	0.95	1.00
2.0	α_1	100	20000	0	0.000154736	0.000003761	0.68	0.95	0.99
2.0	α_2	100	20000	0.51	0.00029991	0.00008595	0.65	0.92	0.99
5.0	α_0	100	20000	0	0.000993973	0.000127154	0.72	0.94	1.00
5.0	α_1	100	20000	0	0.000014310	0.000000296	0.64	0.90	0.99
5.0	α_2	100	20000	0.61	0.00052970	0.00008346	0.66	0.95	0.98
7.0	α_0	100	20000	0	−0.000922645	0.000211488	0.61	0.93	1.00
7.0	α_1	100	20000	0	−0.000039807	0.000000043	0.65	0.91	0.95
7.0	α_2	100	20000	0.61	0.001339981	0.000083268	0.64	0.96	0.99

TABLE 2. $M/M/\infty$ Queue-Length Process with $\lambda = 2.0$, $\mu = 1.0$

t	Estimator	Replications	c	r	Error	MSE	1σ	2σ	3σ
0.5	α_0	100	20000	0	0.000988069	0.000020286	0.64	0.97	1.00
0.5	α_1	100	20000	0	0.000220167	0.000085861	0.65	0.93	1.00
0.5	α_2	100	20000	0.29	−0.00200724	0.00018906	0.79	0.97	0.99
1.0	α_0	100	20000	0	−0.000526618	0.000072227	0.67	0.91	0.99
1.0	α_1	100	20000	0	0.003797064	0.000086822	0.68	0.96	1.00
1.0	α_2	100	20000	0.57	0.002258368	0.000299082	0.74	0.99	1.00
2.0	α_0	100	20000	0	0.000167566	0.000132771	0.74	0.99	1.00
2.0	α_1	100	20000	0	0.002803725	0.000073235	0.64	0.91	1.00
2.0	α_2	100	20000	0.46	−0.003322026	0.000462243	0.65	0.92	1.00
5.0	α_0	100	20000	0	−0.001049106	0.000409169	0.73	0.95	1.00
5.0	α_1	100	20000	0	0.000505925	0.000004452	0.56	0.91	0.97
5.0	α_2	100	20000	0.70	−0.007038762	0.000381976	0.60	0.90	0.98
7.0	α_0	100	20000	0	−0.005216124	0.000756011	0.68	0.97	0.99
7.0	α_1	100	20000	0	0.000116769	0.000000512	0.61	0.86	0.94
7.0	α_2	100	20000	0.71	−0.004858809	0.000265057	0.70	0.95	0.99

$$P_{ij} = \begin{cases} q & \text{if } j = i+1 \\ 0 & \text{if } j > i+1 \\ p^{i-j+1}q & \text{if } i \geq j \geq 1 \\ p^{i+1} & \text{if } j = 0. \end{cases}$$

Now we generate $EW(n)$ iteratively. Let δ_j^N denote the probability that when the Nth customer arrives in the system, there will already be j customers present. Setting $\delta_0^0 = 1$, the vector $\{\delta_j^N : j \geq 0\}$ may be computed recursively in N. Then, the expected waiting time for the Nth arriving customer is given by

$$E(W(N)) = \frac{1}{\mu} \sum_{j=1}^{N-1} j\delta_j^N.$$

The experimental results for this model can be found in Tables 3 and 4.

Again the performance (measured in mean square error) of the conventional estimator degrades as $t \to \infty$. The coupling-based estimator continues to have superior performance for large t. The empirically based coupling estimator $\alpha_2(c)$ dominates the conventional estimator as $t \to \infty$, since it does not change significantly as t gets large. Note that the confidence interval coverages for the coupling-based estimator $\alpha_1(c)$ are not close to their nominal values. The reason is that the tiny variances contribute to a relatively large skewness; this creates small-sample difficulties in the normal approximation.

4. PROOFS

PROOF OF PROPOSITION 1: For any $x \in S$,

$$\begin{aligned} &\|P(X(t) \in \cdot | X(0) = x) - P(X(t) \in \cdot)\| \\ &\quad = \|P_x(X_2(t) \in \cdot) - P_x(X_1(t) \in \cdot)\| \leq P_x(T > t) \to 0 \end{aligned}$$

as $t \to \infty$. On the other hand,

$$\|\pi(\cdot) - P(X(t) \in \cdot)\| \leq \int_S \pi(dx) \|P(X(t) \in \cdot | X(0) = x) - P(X(t) \in \cdot)\|,$$

so $\|\pi(\cdot) - P(X(t) \in \cdot)\| \to 0$ as $t \to \infty$ also. Hence, Eq. (2.3) follows. Furthermore, for any bounded f, Eq. (2.3) implies that $E[\, f(X(t)) | X(0) = x] \to Ef(X^*(0))$ as $t \to \infty$. We may then invoke Theorem 1 of Glynn [6] to conclude that X is Harris recurrent. ■

PROOF OF THEOREM 1: Let

$$R_i(c) = \int_{[0, T_i(c))} [\, f(X_{i1}(c,t)) - f(X_{i2}(c,t))] G(dt).$$

TABLE 3. $M/M/1$ Waiting Time Process with $\lambda = 0.2$, $\mu = 1.0$

N	Estimator	Replications	c	r	Error	MSE	1σ	2σ	3σ
1.0	α_0	100	20000	0	−0.00027	0.00002	0.64	0.94	1.00
1.0	α_1	100	20000	0	0.00352	0.00005	0.46	0.81	0.95
1.0	α_2	100	20000	0.10	−0.00044	0.00035	0.63	0.94	0.98
2.0	α_0	100	20000	0	−0.00066	0.00005	0.64	0.94	1.00
2.0	α_1	100	20000	0	0.00322	0.00004	0.48	0.77	0.93
2.0	α_2	100	20000	0.42	−0.00113	0.00015	0.66	0.96	0.99
5.0	α_0	100	20000	0	0.00084	0.00016	0.66	0.93	1.00
5.0	α_1	100	20000	0	0.00051	0.00001	0.52	0.80	0.98
5.0	α_2	100	20000	0.47	−0.00110	0.00014	0.72	0.90	0.97
7.0	α_0	100	20000	0	−0.00154	0.00019	0.65	0.96	1.00
7.0	α_1	100	20000	0	−0.00001	0.00000	0.49	0.93	0.97
7.0	α_2	100	20000	0.58	0.00084	0.00016	0.63	0.92	0.98
10.0	α_0	100	20000	0	0.00089	0.00035	0.63	0.93	0.99
10.0	α_1	100	20000	0	0.00005	0.00000	0.23	0.67	0.79
10.0	α_2	100	20000	0.60	−0.00174	0.00014	0.55	0.88	0.98

TABLE 4. $M/M/1$ Waiting Time Process with $\lambda = 0.5$, $\mu = 1.0$

N	Estimator	Replications	c	r	Error	MSE	1σ	2σ	3σ
5.0	α_0	100	20000	0	0.00047	0.00051	0.66	0.94	0.99
5.0	α_1	100	20000	0	0.04128	0.00240	0.17	0.41	0.70
5.0	α_2	100	20000	0.12	0.00321	0.00513	0.58	0.95	1.00
7.0	α_0	100	20000	0	−0.00284	0.00087	0.63	0.97	1.00
7.0	α_1	100	20000	0	0.03006	0.00146	0.29	0.60	0.86
7.0	α_2	100	20000	0.54	−0.00820	0.00264	0.75	0.92	1.00
10.0	α_0	100	20000	0	0.00129	0.00136	0.66	0.97	0.99
10.0	α_1	100	20000	0	0.01532	0.00066	0.40	0.78	0.93
10.0	α_2	100	20000	0.59	−0.00473	0.00238	0.69	0.93	0.99
20.0	α_0	100	20000	0	0.00555	0.00361	0.63	0.95	0.98
20.0	α_1	100	20000	0	0.00497	0.00022	0.49	0.79	0.91
20.0	α_2	100	20000	0.72	0.00554	0.00240	0.68	0.95	0.99
50.0	α_0	100	20000	0	−0.01112	0.00724	0.66	0.97	0.99
50.0	α_1	100	20000	0	−0.00007	0.00001	0.2	0.6	0.68
50.0	α_2	100	20000	0.53	0.00647	0.00282	0.64	0.97	1.00

Note that

$$\begin{aligned}&\|P(R_1(c) \in \cdot | X) - P^*(R \in \cdot)\| \\ &\quad = \left\| \int_S (\pi_c(dx) - \pi(dx)) P_x(R \in \cdot) \right\| \leq \|\pi_c - \pi\| \to 0 \text{ a.s.}\end{aligned}$$

as $c \to \infty$. Furthermore,

$$\begin{aligned}E(R_1^2(c)|X) &= \int_S \pi_c(dx) E_x R^2 \\ &= \int_S \pi_c(dx) g_2(x) \to \int_S \pi(dx) g_2(x) = E^* R^2 \text{ a.s.}\end{aligned}$$

as $c \to \infty$, so $\{R_1^2(c) : c > 0\}$ is a uniformly integrable family of r.v.'s. As a consequence, we may apply the Lindeberg–Feller CLT (see Chung [2, p. 205]) path-by-path to conclude that

$$P\left(\frac{1}{\sqrt{\lambda c_2}} \sum_{i=1}^{\lfloor \lambda c_2 \rfloor} (R_i(c) - E(R_1(c)|X)) \leq \cdot | X \right) \Rightarrow P(\sigma_2 N(0,1) \leq \cdot) \text{ a.s.} \qquad \textbf{(4.8)}$$

as $c \to \infty$. Also, for $\epsilon > 0$,

$$\begin{aligned}P(N(c_2)) < \lambda c_2(1-\epsilon)|X) &\leq P\left(\sum_{i=1}^{\lceil \lambda c_2(1-\epsilon) \rceil} \chi_i(c) > c_2 | X \right) \\ &\leq P\Bigg(\sum_{i=1}^{\lceil \lambda c_2(1-\epsilon) \rceil} (\chi_i(c) - E[\chi_1(c)|X]) \\ &\qquad > c_2 - \lceil \lambda c_2(1-\epsilon) \rceil E[\chi_1(c)|X] | X \Bigg) \\ &\leq \frac{\text{var}(\chi_1(c)|X) \lceil \lambda c_2(1-\epsilon) \rceil}{(c_2 - \lceil \lambda c_2(1-\epsilon) \rceil E[\chi_1(c)|X])^2} \to 0 \text{ a.s.}\end{aligned}$$

as $c \to \infty$, by virtue of A3 and A5. Similarly, $P(N(c_2)) > \lambda c_2(1+\epsilon)|X) \to 0$ a.s. as $c \to \infty$. So,

$$P(|N(c_2) - \lambda c_2| > \epsilon c_2 | X) \to 0 \text{ a.s.} \qquad \textbf{(4.9)}$$

as $c_2 \to \infty$. Furthermore, because $R_1(c), R_2(c), \ldots$ are independent conditional on X, Kolmogorov's Inequality implies that for $\epsilon > 0$,

$$\begin{aligned}&P\left(\left| \sum_{i=1}^{\lfloor \lambda c_2 \rfloor} (R_i(c) - E(R_i(c)|X)) - \sum_{i=1}^{N(c_2)} (R_i(c) - E(R_i(c)|X)) \right| > \epsilon \sqrt{c_2} \;\middle|\; X \right) \\ &\quad \leq P\left(\max_{1 \leq |j| \leq \epsilon^3 c_2} \left| \sum_{i=\lfloor \lambda c \rfloor}^{\lfloor \lambda c \rfloor + j} (R_i(c) - E(R_i(c)|X)) \right| > \epsilon \sqrt{c_2} \;\middle|\; X \right) \\ &\qquad + P(|N(c_2) - \lambda c_2| > \epsilon^3 c_2 | X) \\ &\quad \leq \frac{\text{var}(R_1(c)|X) \epsilon^3 c_2}{\epsilon^2 c_2} + P(|N(c_2) - \lambda c_2| > \epsilon^3 c_2 | X).\end{aligned}$$

If we let $c \to \infty$, followed by letting $\epsilon \to 0$ (and apply A3, A4, Eq. (4.1), and Eq. (4.2)), we may conclude that

$$P\left(\frac{1}{\sqrt{\lambda c_2}}\sum_{i=1}^{N(c_2)}(R_i(c)-E[R_i(c)|X]) \le \cdot |X\right) \Rightarrow P(\sigma_2 N(0,1) \le \cdot) \text{ a.s.}$$

as $c \to \infty$. Utilizing Eq. (4.2) again, we find that

$$P\left(\sqrt{c}\left(\frac{1}{N(c_2)}\sum_{i=1}^{N(c_2)}(R_i(c)-E[R_i(c)|X])\right) \le \cdot |X\right)$$
$$\Rightarrow P\left(\frac{\sigma_2}{\sqrt{(1-r)\lambda}}N(0,1) \le \cdot\right) \text{ a.s.}$$

as $c \to \infty$. Hence, for each θ,

$$E\exp\left(i\theta\sqrt{c}\left(\frac{1}{N(c_2)}\sum_{j=1}^{N(c_2)}(R_j(c)-E[R_1(c)|X])\right)|X\right) \to \exp\left(\frac{-\theta^2\sigma_2^2}{2(\lambda(1-r))}\right)$$

as $c \to \infty$. Consequently, noting that

$$\Gamma_j(c) = R_j(c) - E[R_1(c)|X] + \int_S \pi_c(dx)k(x),$$

we get

$$\begin{aligned}
&E\exp(i\theta\sqrt{c}(\alpha(c)-\alpha)) \\
&\quad = E\exp\left(i\theta\sqrt{c}\left(\frac{1}{N(c_2)}\sum_{j=1}^{N(c_2)}(R_j(c)-E[R_1(c)|X]) + \int_S \pi_c(dx)k(x) - \alpha\right)\right) \\
&\quad = E\left\{\exp\left(i\theta\sqrt{c}\left(\frac{1}{c_1}\int_0^{c_1} k(x)\pi_c(dx) - \alpha\right)\right)\right. \\
&\qquad \left.\times E\exp\left(i\theta\sqrt{c}\left(\frac{1}{N(c_2)}\sum_{j=1}^{N(c_2)}(R_j(c)-E[R_1(c)|X])\right)\right)\right\} \\
&\quad \to \exp\left(-\frac{\sigma_1^2\theta^2}{2r} - \frac{\sigma_2^2\theta^2}{2\lambda(1-r)}\right)
\end{aligned}$$

as $c \to \infty$, proving the theorem. ■

PROOF OF THEOREM 2: Using A7, a proof very similar to that of Theorem 1 establishes that

$$\sqrt{c}(\alpha_1(c)-\alpha,\ldots,\alpha_m(c)-\alpha) \Rightarrow \sigma\sqrt{m}(N_1(0,1),\ldots,N_m(0,1))$$

as $c \to \infty$, where $N_1(0,1),\ldots,N_m(0,1)$ are i.i.d. normally distributed r.v.'s with mean zero and unit variance. The conclusion follows from an application of the continuous mapping principle (see Billingsley [1]). ■

PROOF OF PROPOSITION 2: Note that

$$\begin{aligned}
&\sup_B \left| \frac{1}{c_1} \int_0^{c_1} I(X(s) \in B)\, ds - \pi(B) \right| \\
&\quad = \sup_B \left| \sum_{y \in B} \frac{1}{c_1} \int_0^{c_1} I(X(s) = y)\, ds - \pi(y) \right| \\
&\quad \le \sum_{y \in S} \left| \frac{1}{c_1} \int_0^{c_1} I(X(s) = y)\, ds - \pi(y) \right| \\
&\quad = 2 \sum_{y \in S} \left(\pi(y) - \frac{1}{c_1} \int_0^{c_1} I(X(s) = y)\, ds \right) I\left(\pi(y) > \frac{1}{c_1} \int_0^{c_1} I(X(s) = y)\, ds \right).
\end{aligned}$$

The summands are dominated by $2\pi(y)$ (which is summable) and converge to zero a.s. Applying the dominated convergence theorem path-by-path leads to the conclusion that the above sum converges to zero a.s. as $c \to \infty$. ■

PROOF OF PROPOSITION 3: Let

$$p(y) = \int_S \pi(dx) p(x,y),$$

$$p_c(y) = \frac{1}{c_1} \int_0^{c_1} p(X(s), y)\, ds.$$

By stationarity of π,

$$\begin{aligned}
\pi(B) &= \int_S \pi(dx) \int_B p(x,y) \eta(dy) \\
&= \int_B p(y) \eta(dy).
\end{aligned}$$

Also,

$$\pi_c(B) = \int_B p_c(y) \eta(dy).$$

Fix $\epsilon > 0$. Because η and π are tight [1], there exists a compact set K such that $\eta(K^c) < \epsilon$, and $\pi(K^c) < \epsilon$. Set $\|p\| = \sup\{p(x,y) : x, y \in S\}$. Note that $\|p\| \ge p(y)$, $\|p\| \ge p_c(y)$ for $c > 0$, $y \in S$. Since p is necessarily uniformly continuous on $K \times K$, there exists $y_1, y_2, \ldots, y_l \in K$ such that for each $y \in K$, there is a y_i for which

$$|p(x,y) - p(x,y_i)| < \epsilon$$

for $x \in K$. Hence,

$$\begin{aligned}
&\sup_{y\in K}|p_c(y) - p(y)| \\
&\quad \le \sup_{y\in K}\left|\int_K p(x,y)\left(\frac{1}{c_1}\int_0^{c_1} I(X(s)\in dx)\,ds - \pi(dx)\right)\right| + \epsilon\|p\| \\
&\quad \le \max_{1\le i\le l}\left|\int_K p(x,y_i)\left(\frac{1}{c_1}\int_0^{c_1} I(X(s)\in dx)\,ds - \pi(dx)\right)\right| + 3\epsilon\|p\| \\
&\quad = \max_{1\le i\le l}\left|\frac{1}{c_1}\int_0^{c_1} p(X(s),y_i)I(X(s)\in K)\,ds\right. \\
&\qquad\qquad \left. - \int_S p(x,y_i)I(x\in K)\pi(dx)\right| + 3\epsilon\|p\| \to 3\epsilon\|p\| \text{ a.s.} \qquad \textbf{(4.10)}
\end{aligned}$$

as $c \to \infty$, by the law of large numbers for Harris processes (applied to the finite collection of sample functions associated with $y_1, \ldots, y_l$). Then,

$$\begin{aligned}
\sup_B|\pi_c(B) - \pi(B)| &= \sup_B\left|\int_B p_c(y)\eta(dy) - \int_B p(y)\eta(dy)\right| \\
&\le \int_S |p_c(y) - p(y)|\eta(dy) \\
&\le \int_K |p_c(y) - p(y)|\eta(dy) + \epsilon\|p\| \\
&\le \sup_{y\in K}|p_c(y) - p(y)| + \epsilon\|p\|.
\end{aligned}$$

By virtue of Eq. (4.10), we conclude that

$$\overline{\lim_{c\to\infty}} \sup_B|\pi_c(B) - \pi(B)| \le 4\epsilon\|p\|.$$

Since ϵ was arbitrary, this proves the result. ■

References

1. Billingsley, P. (1968). *Convergence of probability measure.* New York: John Wiley.
2. Chung, K.L. (1974). *A course in probability theory.* New York: Academic Press.
3. Devroye, L. (1985). *Nonparametric density estimation: The L1 view.* New York: John Wiley.
4. Fox, B.L. & Glynn, P.W. (1989). Estimating discounted costs. *Management Science* 35: 1297–1325.
5. Glynn, P.W. (1989). A GSMP formalism for discrete-event systems. *Proceedings of the IEEE* 77: 14–23.
6. Glynn, P.W. (1994). Some topics in regenerative steady state simulation. *Acta Applicandae Mathematicae* 34: 225–236.
7. Glynn, P.W. & Wong, E.W. (1996). Efficient simulation via coupling. *Probability in the Engineering and Informational Sciences* 10: 165–186.
8. Lindvall, T. (1992). *Lectures on the coupling method.* New York: John Wiley.