PROBABILITY IN THE ENGINEERING AND INFORMATIONAL SCIENCES

Editor
Sheldon Ross
University of California, Berkeley

Editorial Office: *Probability in the Engineering and Informational Sciences*, College of Engineering, Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720, U.S.A. Tel. (510) 642-3615.

Publishing, Production and Advertising Offices: Cambridge University Press, 40 West 20th Street, New York, NY 10011, U.S.A.

Subscription Offices: Cambridge University Press, 110 Midland Avenue, Port Chester, NY 10573, U.S.A. (for US/Canada); or Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU, England (for UK/rest of world).

1996 Subscription Information: *Probability in the Engineering and Informational Sciences* (ISSN 0269-9648) is published quarterly. Institutional rate: US $274.00 in the U.S.A. and Canada; UK £185.00 in all other countries. Individual rates (prepayment by personal check or credit card only): US $80.00 in the U.S.A. and Canada, UK £59.00 in all other countries. Prices include postage and insurance; air mail UK £16.00 per year extra.

Article Reprints are available from the authors. For orders of 50 or more reprints, contact the Journals Production Office.

Application to mail at second-class rates is pending in New York, NY, and additional mailing offices. Postmaster: Send address changes in the U.S.A. and Canada to *Probability in the Engineering and Informational Sciences*, Subscription Department, Cambridge University Press, 110 Midland Avenue, Port Chester, NY 10573.

Printed in the United States of America.

EFFICIENT SIMULATION VIA COUPLING

PETER W. GLYNN* AND EUGENE W. WONG
Department of Operations Research
Stanford University
Stanford, California 94305-4022

This paper is concerned with how coupling can be used to enhance the efficiency of a certain class of terminating simulations, in Markov process settings in which the stationary distribution is known. We are able to theoretically establish that our coupling-based estimator is often more efficient than the naive estimator. In addition, we discuss extensions of our methodology to Markov process settings in which conventional coupling fails and show (for Doeblin chains) that knowledge of the stationary distribution is sometimes unnecessary.

1. INTRODUCTION

There exists a substantial number of stochastic processes for which a great deal is known about the steady-state behavior but for which the transient behavior is analytically intractable. Among the systems that have this property are product-form Markovian queueing networks (see, e.g., Kelly [15]), BCMP networks [6], loss networks [16], and models of polymerization and random fields [22]. Other such examples are scattered throughout the applied probability literature.

In this paper, we focus on how such steady-state information can be used to enhance the efficiency of a certain class of terminating simulations. The main idea that we shall exploit is that of coupling. This concept has had tremendous

*This research was supported by the National Science Foundation under grant DDM-9101580 and the Army Research Office under contract DAAL03-91-G-0319.

0269-9648/96 $11.00 + .10

impact on Markov process theory over the last 20 years. It involves running a stationary version of a Markov process, concurrently with the nonstationary version, until the two processes meet (or *couple*). As might be expected, the magnitude of the coupling time yields significant theoretical information about the rate at which the nonstationary version of the Markov process converges to its steady-state (see, e.g., Lindvall [18]). This has also been exploited, in a simulation context, to obtain a computational algorithm for estimating upper bounds on the rate of convergence to stationarity (see Kalashnikov [14]).

In Section 2 of this paper, we describe how coupling can be used to obtain a new class of estimators that take full advantage of knowledge of the stationary distribution. We then provide some simple examples of how such couplings can be constructed for discrete-time and continuous-time Markov chains.

Section 3 is concerned with analysis of the efficiency of this new class of estimators. We show that when coupling is applied the computation of $\alpha = Ef(X(t))$ (for $X = \{X(t): t \geq 0\}$ a Markov process) actually becomes easier as $t \to \infty$. This contrasts with the degradation of the naive estimator's efficiency as $t \to \infty$ (because X must then be simulated to time t along each replication, making replications more expensive as $t \to \infty$). We are similarly able to establish the superiority of our coupling-based estimator for estimation of expected cumulative cost over intervals $[0, t]$ with t large.

In Section 4, we discuss the construction of couplings in the context of Harris chains and describe two different implementations. We also give an extension of our methodology to periodic chains; these are chains to which conventional coupling does not apply. We close the section by showing that for the class of Doeblin chains a modified version of our proposed algorithm can be applied without any knowledge of the stationary distribution. Although not as efficient as our original coupling-based estimator, the modified estimator continues to be more efficient than the naive estimator for estimating $Ef(X(t))$ with t large.

For many processes, and particularly those in continuous time, the construction of a coupling is problematic, both from a theoretical standpoint and from a programming standpoint. However, a concept known as *shift-coupling* (see Aldous and Thorisson [2]) substantially weakens the requirements demanded of the coupling and is, in general, much easier to implement computationally. A major drawback of the shift-coupling methodology described in Section 5 is that it applies only to performance measures that can be expressed as an expected cumulative cost.

Section 6 is devoted to how knowledge of the stationary distribution can be exploited in Markov process settings in which both coupling and shift-coupling are inapplicable. We describe an approach, based on control variates, that asymptotically dominates naive sampling from an efficiency viewpoint. Finally, in Section 7, we provide a brief account of our computational experience with our algorithms.

2. THE BASIC IDEA

Let $X = \{X(t): t \geq 0\}$ be a continuous-time Markov process living on state-space S. (To handle a chain $\{X_n: n \geq 0\}$, we embed the process in continuous time by setting $X(t) = X_{\lfloor t \rfloor}$, where $\lfloor t \rfloor$ is the greatest integer less than or equal to t.) Our goal is to compute, via simulation, an expectation of the form

$$\alpha = \int_{[0,\infty)} Ef(X(s))G(ds), \tag{2.1}$$

where $f: S \to \Re$ is a real-valued performance measure and $G(.)$ is a deterministic finite measure on $[0,\infty)$.

Example 1: The problem of estimating the transient quantity $\alpha = Ef(X(t))$ is a special case of the preceding, as can be seen by setting $G(ds) = \delta_t(ds)$. (δ_t is a point mass probability concentrated at t.)

Example 2: If $f(x)$ is interpreted as the rate at which "cost" accrues while X occupies state x, then

$$\alpha = E\int_0^t f(X(s))\, ds \tag{2.2}$$

is the expected cumulative cost incurred over the interval $[0, t]$. Note that if we set $G(ds) = I(0 \leq s \leq t)\, ds$, we can incorporate Eq. (2.2) into the framework associated with Eq. (2.1).

Example 3: For $\gamma > 0$, consider the infinite-horizon discounted cost given by

$$\alpha = E\int_{[0,\infty)} e^{-\gamma s} f(X(s))\, ds. \tag{2.3}$$

Here we can set $G(ds) = e^{-\gamma s}\, ds$ to fit the problem into the setting of Eq. (2.1).

To apply our coupling idea to the estimation of the expected value of Eq. (2.1), we need to make several assumptions about X:

A1. $EY' < \infty$, where $Y' \stackrel{\text{def}}{=} \int_{[0,\infty)} |f(X(s))|\, G(ds)$.

A2. X possesses a stationary distribution π such that $\nu = \int_S f(x)\pi(dx)$ can be computed.

A3. We can simulate a process $\{(X'(t), X^*(t)): t \geq 0\}$ such that

1. $X'(t) \stackrel{\text{D}}{=} X(t)$ for $t \geq 0$, where $\stackrel{\text{D}}{=}$ denotes equality in distribution,
2. $\{X^*(t): t \geq 0\}$ is a stationary version of X under which $X(0)$ has distribution π, and
3. there exists a finite-valued random time T such that $X'(t) = X^*(t)$ for $t \geq T$.

The random time T hypothesized in Assumption A3.3 is known, in the literature, as a *coupling time* (see, e.g., Lindvall [18]).

We will now show how the preceding assumptions can be fruitfully exploited to produce coupling-based simulation estimators for α. Assumption A1 guarantees that the estimation problem is well defined, whereas Assumption A3.1 permits us to substitute the process X' for X, yielding

$$\alpha = \int_{[0,\infty)} Ef(X'(s))G(ds). \tag{2.4}$$

On the basis of Assumption A3.3, we find that

$$\begin{aligned} Ef(X'(s)) &= E(f(X'(s)); T \le s) + E(f(X'(s)); T > s) \\ &= E(f(X^*(s)); T \le s) + E(f(X'(s)); T > s) \\ &= Ef(X^*(s)) + E(f(X'(s)) - f(X^*(s)); T > s). \end{aligned} \tag{2.5}$$

But Assumptions A3.2 and A2 ensure that $Ef(X^*(s)) = \nu$ is known. Hence, we may rewrite α in the form $\alpha = EY_1$, where

$$Y_1 = \beta + \int_{[0,\infty)} (f(X'(s)) - f(X^*(s)))G(ds) \tag{2.6}$$

and $\beta \stackrel{\text{def}}{=} \nu \int_{[0,\infty)} G(ds)$. Note that Y_1 is a random variable that is (by assumption) simulatable. The idea is now to estimate α by generating independent and identical replicates of the random variable Y_1.

We now provide some illustrations of how couplings satisfying Assumption A3 may be constructed.

Example 4: Let X be a positive recurrent irreducible continuous-time Markov chain with (unique) stationary distribution π having discrete state-space S, and set $X' = X$. Assume that we know how to generate variates from π. (This will, for example, be feasible for product-form queueing networks.) To construct (X', X^*), simulate X in any way desired, and generate $X^*(0)$ (independent of X) from distribution π. We then let X^* evolve independently until $T = \inf\{t \ge 0 : X^*(t) = X(t)\}$, after which we set $X^*(t) = X(t)$. The coupling time T constructed here, under which X and X^* evolve independently until they meet, is easily shown to be finite almost surely (see Lindvall [17]).

Example 5: Let X be a positive recurrent irreducible aperiodic discrete-time Markov chain, with discrete state-space S. Again, if we assume that $X^*(0)$ can be generated (independently of X) from π, the independent coupling of Example 4 works in this setting. (Note that the construction fails if X is periodic.)

Example 6: Let $W = \{W_n : n \ge 0\}$ be the waiting time sequence of the GI/G/1 queue, and assume that the traffic intensity of the queue is strictly less than unity. Suppose that the stationary distribution π of W is known (as is the case in the M/G/1 and G/M/1 settings). The chain W is an aperiodic Harris chain

with state-space $S = [0,\infty)$ (see Meyn and Tweedie [19]), and the independent coupling again works here. (It turns out that both W and W' will typically be at the origin at the coupling time.)

Although the preceding examples all utilize an independent coupling (in which X' and X^* evolve independently up to T), there exist certain applications in which only dependent couplings can fulfill our conditions. For example, if X is the residual life process associated with a renewal process having a spread-out interrenewal time distribution, then any independent coupling of the type described in Examples 4–6 will never occur in finite time. However, by cleverly creating certain dependencies between X' and X^* prior to T, finite-valued coupling times can be constructed; these "dependent couplings" underlie many recent proofs of the renewal theorem.

It should further be pointed out that dependent couplings can be used to construct coupling times T that occur earlier than those associated with independent coupling, thereby potentially reducing computational effort.

These issues will be addressed in greater detail in Section 4.

3. COMPUTATIONAL EFFORT OF THE COUPLING-BASED ESTIMATOR

Our goal here is to compare the computational efficiency of the coupling-based estimator introduced in Section 2 to that of the naive estimator.

To be more precise, the naive estimator is based on generating i.i.d. replicates $Y(1), Y(2), \ldots$ of the r.v. Y, where

$$Y = \int_{[0,\infty)} f(X(s))G(ds). \tag{3.7}$$

Let $\tau(i)$ be the computer time required to generate $Y(i)$, and assume (reasonably) that the $(Y(i), \tau(i))$'s are i.i.d. If $N(c)$ is the number of $Y(i)$'s produced in c units of computer time, then

$$\alpha(c) = \begin{cases} \dfrac{1}{N(c)} \displaystyle\sum_{i=1}^{N(c)} Y(i) & \text{if } N(c) \ge 1 \\ 0 & \text{otherwise} \end{cases}$$

is the naive estimator for α based on c units of computer time. Glynn and Whitt [12] proved that if $0 < E\tau(1) < \infty$ and $\operatorname{var} Y < \infty$, then

$$c^{1/2}(\alpha(c) - \alpha) \Rightarrow \sigma N(0,1)$$

as $c \to \infty$, where $\sigma^2 = E\tau(1)\operatorname{var} Y$.

A similar analysis can be performed on the coupling-based estimator of Section 2. Let $Y_1(1), Y_1(2), \ldots$ be a sequence of i.i.d. replicates of Y_1, and let $\tau_1(i)$ be the computer time required to generate $Y_1(i)$. If $\alpha_1(c)$ is the corre-

sponding estimator available upon expending c units of computer time, it follows (under appropriate regularity conditions) that

$$c^{1/2}(\alpha_1(c) - \alpha) \Rightarrow \sigma_1 N(0,1)$$

as $c \to \infty$, where $\sigma_1^2 = E\tau_1(1)\text{var } Y_1$. Thus, a comparison of asymptotic efficiency reduces to comparing σ_1^2 to σ^2.

In general, no universal statement about the relative magnitudes of σ_1^2 and σ^2 can be made. This should come as no surprise, given the fact that the magnitude of σ_1^2 depends critically on both the computational and statistical efficiency of the coupling that is chosen. Instead, we shall settle for an asymptotic analysis for σ^2 and σ_1^2.

Example 1 (continued): We are concerned here with a class of estimation problems that clearly depends on the time parameter t. Our notation will now suitably reflect this dependence on t (i.e., $\sigma^2 = \sigma^2(t)$, $Y = Y(t)$, etc.).

To analyze $\sigma^2(t)$, note that it is reasonable to expect $E\tau(1,t)$ to grow linearly in t, so that $E\tau(1,t) \sim ct$ for some $c > 0$. As for var $Y(t)$, assume that X is a positive recurrent irreducible continuous-time Markov chain. Then, we can typically expect that var $Y(t) \to d \stackrel{\text{def}}{=} \text{var } f(X^*(0)) > 0$ as $t \to \infty$.

On the other hand, $Y_1(t)$ requires only that one simulate to time $T \wedge t$. Consequently, it is reasonable to expect that $E\tau_1(1,t) \to c_1 > 0$ as $t \to \infty$. Furthermore, var $Y_1(t) = \text{var } Z(t)$, where $Z(t) = (f(X'(t)) - f(X^*(t)))I(T > t)$. Evidently, var $Z(t) = E[Z^2(t)\,|\,T > t]P(T > t) - (E[Z(t)\,|\,T > t]P(T > t))^2$. In the typical coupling, (X', X^*) jointly evolve as a time-homogeneous Markov process up to the first time T at which (X', X^*) hits the "diagonal" $D = \{(x,x') \in S \times S : x = x'\}$, after which X' and X^* run together. Hence, the coupling time T can be viewed as the "hitting time" of D. The typical behavior for Markov processes is that the tail of such a hitting time is asymptotically exponential, so that there exits $d', \lambda > 0$ such that $P(T > t) \sim d' \exp(-\lambda t)$ as $t \to \infty$ (see, e.g., Kielson [17]). In addition, the same theory provides general conditions under which $P(Z(t) \in \cdot\,|\,T > t) \Rightarrow P(Z^* \in \cdot)$, from which we may conclude that the typical behavior of $\sigma_1^2(t)$ is that $\sigma_1^2(t) \sim c_1 d_1 \exp(-\lambda t)$ for some $d_1 > 0$.

Comparison of $\sigma^2(t)$ and $\sigma_1^2(t)$ suggests that the coupling-based estimator will be (much) superior to the naive estimator, at least for large values of t. This is to be expected because, for large t, $\tau_1(1,t)$ is roughly constant, and $Y_1(t)$ is essentially deterministic.

Example 2 (continued): As in Example 1, this problem has a clear dependence on t and we will again allow our notation to reflect this dependence. Here,

$$Y_1(t) = \beta(t) + \int_0^{T \wedge t} (f(X'(s)) - f(X^*(s)))\, ds$$

and, as in Example 1, it is evident that $E\tau_1(1,t) \to c_1 > 0$. Clearly,

$$Y_1(t) - \beta(t) \to \int_0^T (f(X'(s)) - f(X^*(s)))\, ds$$

as $t \to \infty$, so that (under appropriate uniform integrability conditions) var $Y_1(t) \to d_1 > 0$ as $t \to \infty$. On the other hand, $Y(t)$ typically satisfies a central limit theorem (CLT) of the form

$$t^{-1/2}(Y(t) - tEf(X^*(0))) \Rightarrow dN(0,1)$$

as $t \to \infty$. Again, under appropriate uniform integrability conditions, it follows that t^{-1} var $Y(t) \to d^2$ as $t \to \infty$, so that we may conclude that $\sigma^2(t) \sim cd^2t^2$ as $t \to \infty$, whereas $\sigma_1^2(t) \to c_1 d_1$ as $t \to \infty$. Thus, the coupling-based estimator is again (much) superior to the naive estimator, for large t.

In Example 3, it is clear that the computational effort required to compute Y via naive simulation is infinite, whereas the time required to compute

$$Y_1 = \int_0^T e^{-\gamma s}(f(X'(s)) - f(X^*(s)))\, ds$$

is, of course, finite. In particular, the naive estimator is infeasible to compute exactly in finite time, so that the coupling-based estimator Y_1 again dominates. (A complete analysis of some alternatives to the naive estimator, including finite-horizon truncations of the naive estimator, are given in Fox and Glynn [9].)

As indicated earlier, the nature of the coupling used can have a significant impact on the relative efficiency of estimators based on Y_1 versus those based on Y. For example, if the independent coupling of Section 2 is employed, the effort required to simulate the joint process (X', X^*) up to the coupling time T is essentially twice that required to simulate X up to time T. (Of course, Y_1 requires no additional simulation beyond T, whereas Y may require such additional simulation.)

An example of an alternative coupling will now be provided, in the setting of a positive recurrent irreducible continuous-time Markov chain. Let Z be a random variable having the stationary distribution π, distributed independently of X. Set

$$S = \inf\{t \geq 0 : X(t) = Z\},$$

and note that $X^*(t) \stackrel{\text{def}}{=} X(S + t)$ is a stationary version of X. Hence, if we set

$$T = \inf\{t \geq 0 : X(t) = X(S + t)\},$$

T is a coupling time that requires simulation of only one process, namely X, up to T (actually, to $T + S$). (Of course, the coupling time T just proposed may be much larger, on average, than that associated with the independent coupling.)

A concept closely related to that of coupling is that of "strong uniform times," introduced by Aldous and Diaconis [1]. Specifically, let X be a positive recurrent irreducible Markov chain living on a discrete state-space S and suppose that T is a finite-valued stopping time satisfying

$$P(X(t) \in \cdot \,|\, T = t) = \pi(\cdot)$$

for all $t \geq 0$; such a random time T is called a strong uniform time. Prescriptions for constructing such random times exist for certain Markov chains with special structure.

Given such a random time T, it is easily shown that $\alpha = \int_{[0,\infty)} Ef(X(s))G(ds)$ can be expressed as

$$\alpha = \beta + E\int_{[0,T)} [f(X(s)) - \nu]G(ds).$$

One can then construct estimators for α analogous to that described earlier in the coupling context. Note, however, that the computational efficiency comparison in this setting is often even more favorable than that obtained in the coupling context, because of the fact that no stationary version X^* need be simulated, thereby reducing the computational effort. On the other hand, construction of strong uniform times appears to demand that more structure be present in X than that required for coupling times.

4. EXTENSIONS TO HARRIS CHAINS

As mentioned in Section 2, dependent couplings are needed for certain Markov process applications. For example, consider a Markov chain, taking values in a general state space, that visits no point infinitely often. (There exist such processes for which stationary distributions can be computed, as evidenced, for example, by the skeleton chain associated with the residual life process.) Here, it is clear that the associated independent coupling will not couple in finite time with probability 1.

Consequently, dependent couplings must be constructed for such chains. We will focus here on the class of aperiodic, positive recurrent Harris chains; this turns out to be precisely the class of chains for which finite-valued couplings exist for all possible initial distributions (see Lindvall [18, p. 102]).

Recall that $\{X_n : n \geq 0\}$ is said to be a *Harris chain* if there exists a nonnegative function $\lambda(\cdot)$, $\epsilon > 0$, $m \geq 1$, and a probability distribution φ on S such that

$$P(\lambda(X_n) \geq \epsilon \text{ infinitely often} \,|\, X_0 = x) = 1, \tag{4.8}$$

for all $x \in S$;

$$P(X_m \in \cdot \,|\, X_0 = x) \geq \lambda(x)\varphi(\cdot), \tag{4.9}$$

for all $x \in S$. The crucial observation relating to the preceding is that Eq. (4.9) permits us to write the m-step transition probabilities in the form

$$P(X_m \in \cdot \,|\, X_0 = x) = \lambda(x)\varphi(\cdot) + (1 - \lambda(x))Q(x, \cdot), \tag{4.10}$$

where $Q(x, \cdot)$ is a probability distribution on S for each $x \in S$. (This is due to Athreya and Ney [5] and Nummelin [20].) Relation (4.10) can be interpreted probabilistically, as follows. When the chain is in state x, we flip a coin. If the coin flip is successful (with probability $\lambda(x)$), we distribute X according to φ m time units later. Otherwise, we distribute X according to $Q(x, \cdot)$ m time units later. (The intermediate values of the chain are then filled in by conditioning on the initial value x and the "final state" corresponding to the chain m time units later.)

This randomization, based on coin flips, can be used to construct a dependent coupling for aperiodic, positive recurrent Harris chains. We initialize X' and X^* independently. With probability $\lambda(X_0^*) \wedge \lambda(X_0')$, we force both chains to simultaneously distribute themselves according to φ m time units later. Specifically, we generate a point from distribution φ and send both chains simultaneously to that point; this, of course, constitutes a coupling time for (X', X^*).

We now discuss what happens if coupling does not occur. With probability $\lambda(X_0^*) - (\lambda(X_0^*) \wedge \lambda(X_0'))(\lambda(X_0') - (\lambda(X_0^*) \wedge \lambda(X_0')))$, $X^*(X')$ distributes itself according to φ m time units later, whereas $X'(X^*)$ distributes itself according to $Q(X_0', \cdot)(Q(X_0^*, \cdot))$. Finally, with probability $1 - (\lambda(X_0^*) \vee \lambda(X_0'))$, we distribute $X_m'(X_m^*)$ according to $Q(X_0', \cdot)(Q(X_0^*, \cdot))$.

This process is then repeated at the time epochs $m, 2m, 3m, \ldots$. Because of Eq. (4.8) and the aperiodicity, successful coupling is then guaranteed to occur in finite time.

One problem with the preceding approach is that it requires the simulationist to write code to generate variates from the probability distributions $Q(x, \cdot), x \in S$. In addition, when $m > 1$, additional code for generating the "intermediate values" from the appropriate conditional distributions needs to be produced. We now describe a method of producing a coupling that avoids this difficulty.

Note that Eq. (4.9) guarantees that, whenever $\lambda > 0$, φ is absolutely continuous with respect to $P(X_m \in \cdot \,|\, X_0 = x)$. The Radon–Nikodym theroem then ensures that there exists a function w such that

$$\varphi(dy) = w(x, y)P(X_m \in dy \,|\, X_0 = x).$$

The function w necessarily takes values almost surely in $[0,1]$. The idea now is that, rather than explicitly generating variates from $Q(x, \cdot)$, we implement this via acceptance–rejection. In particular, suppose that a segment $(X_n, X_{n+1}, \ldots, X_{n+m})$ of the chain is simulated (by any algorithm consistent with the transition probabilities of X). If an independent uniform random variable U is now generated (independently of X), X_{m+n} turns out to be distributed according

to φ if $U \le w(X_n, X_{n+m})$ and according to $Q(X_n, \cdot)$ otherwise. Thus, this acceptance–rejection approach offers an alternative to the method described earlier in this section. (For details, see Glynn and L'Ecuyer [12].)

The goal is now to produce a coupling that takes advantage of this idea. Here, we generate X_0' and X_0^* independently and simulate X' and X^* independently to time m. A uniform random variable U is then generated. If $U \le w(X_0', X_m') \wedge w(X_0^*, X_m^*)$, then both chains simultaneously have distribution φ at time m. Thus, although $X_m' \ne X_m^*$ in general, the distributions are now identical. If U does not satisfy the inequality, this process is repeated at times m, $2m$, $3m, \ldots$ until the uniform r.v. generated does satisfy the appropriate inequality. This algorithm concludes with the construction of a random time T such that $X_T' \overset{D}{=} X_T^*$; this is known, in the literature, as a weak coupling (or distributional coupling). However, it turns out that Eq. (2.5) continues to hold for such weak couplings (see Lindvall [18]), so that estimators based on Y_1 continue to be valid. It should be noted that the main computational obstacle to implementing this algorithm is the need to compute the function w.

In addition, it is easily seen that for this coupling

$$E(Y_1 \mid X', X^*) = \beta + \sum_{j=0}^{\infty} \sum_{k=0}^{m-1} (f(X'_{jm+k}) - f(X^*_{jm+k})) \prod_{l=1}^{j-1} V_l \int_{[jm+k, jm+k+1)} G(ds),$$

where $V_j = 1 - (w(X'_{(j-1)m}, X'_{jm}) \wedge w(X^*_{(j-1)m}, X^*_{jm}))$. In the case that the support of G is bounded, $E(Y_1 \mid X', X^*)$ can be computed in finite time, has a smaller variance than does Y_1, and requires no uniform random variables (but may require more time to compute, on average, than does Y_1).

We now turn to the periodic case. Specifically, we shall now describe how our coupling-based estimator can be extended to periodic Harris chains. Suppose that the period of X is d and that the cyclic classes are $C_1, C_2, \ldots, C_d$. To construct an appropriate coupling in this context, we generate X_0^* from

$$\pi_i(\cdot) = d\pi(\cdot \cap C_i)$$

on $\{X_0' \in C_i\}$ for $1 \le i \le d$ and then use one of the couplings just described in the aperiodic setting. By starting both X_0' and X_0^* in the same cyclic class, it is guaranteed that there will exist a time T at which $X_T \overset{D}{=} X_T^*$. However, Eq. (2.5) no longer holds, in general, because X^* is no longer stationary. The quantity β now needs to be replaced by

$$\beta' = \sum_{i=0}^{\infty} Ef(X_i^*) g_i,$$

where $g_i = \int_{[i, i+1)} G(ds)$. Of course,

$$Ef(X_i^*) = d \sum_{j=1}^{d} \gamma((j+1) \bmod d) P(X_0 \in C_j),$$

where $\gamma(l) = \int_S f(x) I(x \in C_l) \pi(dx)$. To construct an estimator of α based on this coupling requires more information than is the case in the aperiodic setting. In particular, to compute β', one needs to know the $\gamma(l)$'s as well as $P(X_0 \in C_j)$ for $1 \le j \le d$.

We conclude this section with a discussion of aperiodic Doeblin chains. In this context, it turns out that we can construct a variant of our algorithm in which knowledge of ν is not necessary. Such a chain satisfies a uniform version of Eq. (4.9), namely,

$$P(X_m \in \cdot \mid X_0 = x) \ge \lambda \varphi(\cdot) \tag{4.11}$$

for all $x \in S$, where $\lambda > 0$; this type of chain is automatically positive recurrent. Perhaps surprisingly, it is possible here to generate X_0^* without any explicit knowledge of the stationary distribution of π. One approach to generating such as X_0^* would be to exploit an algorithm proposed in Asmussen, Glynn, and Thorisson [4]. However, it has the undesirable property that the expected computer time required to compute X^* is infinite. Instead, let $Q(x, \cdot)$ be defined as in Eq. (4.10) with $\lambda(\cdot) \equiv \lambda$. Then, the following algorithm generates X_0^* in finite expected time.

1. Generate a geometric random variable with parameter λ—call it N.
2. Using initial distribution φ and transition function Q, simulate the Markov chain $(Z_k : 0 \le k \le N)$ up to time N.
3. Generate Z' according to distribution φ.
4. Generate I uniformly on $\{0, 1, \ldots, m-1\}$.
5. Generate X_0^* from $P(X_I \in \cdot \mid X_0 = Z_N, X_m = Z')$.

Thus, even without explicit knowledge of π, we can potentially construct couplings satisfying Assumption A3. Of course, in such a setting, it is unlikely that ν would be known, so that Y_1 cannot be computed. Instead, part of the computational budget must be assigned to estimating β.

Example 1 (continued): Suppose that we partition our computer budget c so that $100p\%$ is assigned to estimating β and $100(1-p)\%$ is assigned to estimating $\xi \triangleq E[(f(X'(t)) - f(X^*(t)))I(T > t)]$. The two quantities are estimated independently of one another. The estimator for β is $\nu(pc)g$, where $\nu(c)$ is the time-average sample mean produced from a simulation of X over the time horizon associated with a budget of c time units and $g \triangleq \int_{[0,\infty)} G(ds)$. The second expectation is estimated by simulating independent replicates of $(f(X'(t)) - f(X^*(t)))I(T > t)$, thereby yielding an estimator $\xi((1-p)c)$. Both estimators typically satisfy CLTs, namely,

$$c^{1/2}(\nu(c) - \nu) \Rightarrow \eta_1 N(0,1),$$

$$c^{1/2}(\xi(c) - \xi) \Rightarrow \eta_2 N(0,1),$$

as $c \to \infty$. Consequently,

$$c^{1/2}(\nu(pc)g + \xi((1-p)c) - \alpha) \Rightarrow \sqrt{\frac{\eta_1^2 g^2}{p} + \frac{\eta_2^2}{1-p}}\, N(0,1) \qquad \textbf{(4.12)}$$

as $c \to \infty$. It is easily calculated that the value p^*, which minimizes the variance of the limiting normal random variable in Eq. (4.12), is

$$p^* = \frac{|\eta_1| g}{|\eta_1| g + |\eta_2|},$$

in which case

$$c^{1/2}(\nu(p^*c) + \xi((1-p^*)c) - \alpha) \Rightarrow (|\eta_1| g + |\eta_2|)N(0,1) \qquad \textbf{(4.13)}$$

as $c \to \infty$.

As in Section 3, it is worth noting that the parameter $\alpha = Ef(X(t))$ is a function of t, so that an asymptotic analysis of the preceding algorithm can potentially be enlightening. First, observe that $|\eta_1|$ is independent of t and that, in this problem, $g \equiv 1$. The behavior of $|\eta_2|$ is precisely that obtained in our analysis of Section 3, namely, $|\eta_2| = |\eta_2(t)| \sim c_1 e^{-\lambda_1 t}$ for some $c_1, \lambda_1 > 0$. Consequently, $(|\eta_1| g + |\eta_2(t)|) \to |\eta_1| g > 0$ as $t \to \infty$, so that the asymptotic variance of this algorithm is insensitive to t. This is to be contrasted to the naive algorithm, which, as pointed out in Section 3, has an asymptotic variance that grows linearly in t. Hence, for large t, the algorithm just proposed for Doeblin chains is more efficient than the naive algorithm.

Example 2 (continued): We now perform a similar analysis on the cumulative cost estimation problem. The structure of the algorithm is precisely that just described in the setting of Example 1, except that now $g = t$ and ξ must be suitably modified. CLT (4.13) continues to hold.

However, the results of an asymptotic analysis are now quite different. As in Section 3, $|\eta_2| = |\eta_2(t)| \to d_1 > 0$ as $t \to \infty$. However, because $g = t$, it is now evident that $(|\eta_1(t)| g + |\eta_2(t)|) \sim |\eta_1| t$ as $t \to \infty$. Consequently, the limiting variance now increases quadratically in t, matching the behavior of the naive estimator for this problem (see Section 3). Hence, there is no clear (large t) advantage to using this algorithm rather than the naive estimator.

It is worth noting that corresponding to every Harris chain there exists a (classically) regenerative sequence with identical marginal distributions and possessing independent (regenerative) cycles (see Glynn [11]). (In general, Harris chains for which the minorization of Eq. (4.9) holds with $m > 1$ have only 1-dependent cycles.) Rather than applying the preceding coupling methods directly to the Harris chain, one could instead apply them to the corresponding classically regenerative sequence. It is for this reason that we included postulate A3.1.

We close this section with a brief comment related to the use of simulation for numerically computing the total variation rate of convergence for a Markov process. It was already shown that in the Doeblin context the simulationist need not know ν in order to exploit the coupling ideas of this paper. Similarly, note that if X is a finite state-space Markov process, our coupling identity of Eq. (2.5) implies that $|P(X(t) \in \cdot) - \pi(\cdot)|$ can be estimated via

$$\sum_{x \in S} \left| \frac{1}{n} \sum_{i=1}^{n} I(X_i'(t) = x, T_i > t) - \frac{1}{n} \sum_{i=1}^{n} I(X_i^*(t) = x, T_i > t) \right|,$$

here $(X_1', X_1^*, T_1), \ldots, (X_n', X_n^*, T_1)$ are i.i.d. replicates of the random triple (X, X^*, T). In particular, explicit knowledge of the stationary distribution is unnecessary; only the ability to generate a stationary version X^* is required. This approach differs from that suggested by Kalashnikov [14], which provided only upper bounds on the total variation distance (but, on the other hand, does not require the ability to generate X^*).

5. ESTIMATION BASED ON SHIFT-COUPLING

An important recent development in the theory of coupling has been the realization that a concept known as "shift-coupling" plays a key role in certain Cesaro-type limit theorems for stochastic processes (see Aldous and Thorisson [2] for details). It turns out that the concept can also play a useful role in simulation. We will now replace Assumption A3 with the following assumption:

A4. We can simulate a process $\{(X'(t), X^*(t)) : t \geq 0\}$ such that

1. $X'(t) \stackrel{D}{=} X^*(t)$ for $t \geq 0$;
2. $\{X^*(t) : t \geq 0\}$ is a stationary version of X under which $X(0)$ has distribution π; and
3. there exist finite-valued random times T and T^* such that $\{X'(T+s) : s \geq 0\} \stackrel{D}{=} \{X^*(T^* + s) : s \geq 0\}$.

The key idea is that we no longer require that $T = T^*$, so that a "shift" of T^* relative to T is now permitted.

One major drawback of the shift-coupling idea that we will now present is that it appears to be appropriate only for cumulative costs, in which case $G(ds) = I(0 \leq s \leq t)\, ds$. Assume Assumptions A1, A2, and A4. Then, for this class of estimation problems, we find that

$$\int_0^t Ef(X(s))\, ds = \int_0^t Ef(X'(s))\, ds$$

$$= E\int_0^T f(X'(s))\, ds + E\int_T^{T+t} f(X'(s))\, ds - E\int_t^{T+t} f(X'(s))\, ds. \qquad \textbf{(5.14)}$$

Similarly, we conclude that

$$\int_0^t Ef(X^*(s))\,ds = E\int_0^{T^*} f(X^*(s))\,ds + E\int_{T^*}^{T^*+t} f(X^*(s))\,ds - E\int_t^{T^*+t} f(X^*(s))\,ds. \tag{5.15}$$

But $\int_0^t Ef(X^*(s))\,ds = \nu t$ by Assumptions A2 and A4.2. Also,

$$E\int_T^{T+t} f(X'(s))\,ds = E\int_{T^*}^{T^*+t} f(X^*(s))\,ds. \tag{5.16}$$

Combining Eqs. (5.14)–(5.16), we obtain the identity $\alpha = EY_2$, where

$$Y_2 = \nu t + \int_0^T f(X'(s))\,ds - \int_0^{T^*} f(X^*(s))\,ds + \int_t^{t+T^*} f(X^*(s))\,ds - \int_t^{t+T} f(X'(s))\,ds. \tag{5.17}$$

Estimation based on Y_2 is often much easier to implement than that based on Y_1. For example, as indicated in Section 2, coupling in continuous time (as exemplified by the residual life process of a renewal process with spread-out interrenewal times) can require fairly complicated dependent couplings that may require substantial effort to develop and program. On the other hand, it is often much easier to shift-couple such processes. This is nicely illustrated by the residual life process; one can independently simulate X and X^* and let T and T^* be the first time that X and X^*, respectively, visit the origin. (Another good example is the queue-length process of the single-server GI/G/1 queue.)

In fact, there exist processes for which couplings, in the sense of Assumption A3, do not exist, whereas shift-couplings do exist. Perhaps the simplest such example is that of a periodic positive recurrent Harris chain. Such chains do not couple in finite time almost surely; however, if one defines T and T^* as the first times at which X and X^*, respectively, possess distribution φ, a shift-coupling is easily obtained. Note that the estimator based on Y_2 then requires significantly less information from the simulationist than that suggested in Section 4, in which the $\gamma(l)$'s and $P(X_0 \in C_i)$'s must be known.

A more subtle example is given by the residual life process associated with interrenewals that are neither spread-out nor arithmetic. The same shift-coupling as described earlier works here; however, it is a consequence of the coupling inequality (and the failure of this residual life process to converge in total variation norm to its steady state) that couplings of the form in Assumption A3 cannot exist here. This difficulty extends, more generally, to (positive recurrent) Harris recurrent Markov processes in continuous time (for definitions and basic properties, see Sigman [21]). Such processes are known to possess shift-couplings but do not, in general, couple as in Assumption A3.

From an implementation standpoint, the shift-coupling can often be constructed without the need to explicitly simulate a stationary version X^*. For example, suppose that X is a discrete state space irreducible, positive recurrent continuous-time Markov chain, and let Z be generated (independently of X) from the stationary distribution π. Set $T = \inf\{t \geq 0 : X(t) = Z\}$, $X^*(s) = X(T+s)$, and $T^* = 0$. A slightly different implementation of this idea is to simulate X^*, generate Z (independently of X^*) from the distribution of $X(0)$, and set $T^* = \inf\{t \geq 0 : X^*(t) = Z\}$, with $X'(s) = X^*(T^* + s)$, $T = 0$. (This second variant works, for example, with the waiting time sequence of Example 6 with $W_0 = 0$ almost surely, whereas the first version fails.)

We conclude this section with an asymptotic (large t) analysis of the efficiency of the shift-coupling estimator introduced in Eq. (5.4). Let $\tau_2(i)$ be the computer time required to generate $Y_2(i)$, where $\{Y_2(i), \tau_2(i)\}$'s are i.i.d. random vectors with $Y_2(i) \stackrel{D}{=} Y_2$. If $\alpha_2(c)$ is the estimator, based on the $Y_2(i)$'s, available after an expenditure of c units of computer time, then (under suitable moment hypotheses)

$$c^{1/2}(\alpha_2(c) - \alpha) \Rightarrow \sigma_2 N(0,1)$$

as $c \to \infty$, where $\sigma_2^2 = E\tau_2(1)\text{var}\, Y_2$. Note that $Y_2(t) - \nu t$ typically converges in distribution as $t \to \infty$, from which it is evident that $\text{var}\, Y_2(t) \to d_2 > 0$ is to be expected. On the other hand, Y_2 requires that the process (X', X^*) be simulated to $t + (T \vee T^*)$, so that the typical behavior of $E\tau_2(1,t)$ is that $E\tau_2(1,t) \sim c_2 t$ as $t \to \infty$, for some $c_2 > 0$. We conclude that $\sigma_2^2(t) \sim c_2 d_2 t$ as $t \to \infty$, so that the shift-coupling estimator proposed in this section is more efficient than the naive estimator for large t but less efficient than the estimator based on coupling.

6. USE OF STEADY-STATE INFORMATION FOR PROCESSES THAT DO NOT ADMIT COUPLING

As noted in Section 5, there exist processes for which couplings do not exist for all possible initial distributions; these include periodic Harris chains and certain Harris recurrent Markov processes in continuous time (such as the residual life process corresponding to interrenewals that are neither spread-out nor arithmetic). In Section 5, the notion of shift-coupling was used to construct estimators for expected cumulative costs that are more efficient (for large time horizons) than those associated with naive sampling.

However, this leaves open the question of whether there exist other ways of taking advantage of steady-state information for such processes that apply to expectations other than expected cumulative costs. In addition, it should be noted that there exist Markov processes to which even the notion of shift-coupling does not apply.

Example 7: Consider a real-valued autoregressive process of order 1, satisfying the recursion

$$X_{n+1} = \tfrac{1}{2}X_n + V_{n+1},$$

where $\{V_n : n \geq 1\}$ is i.i.d. If $P(V_1 = \frac{1}{2}) = P(V_1 = -\frac{1}{2}) = \frac{1}{2}$, then it is straightforward to verify that

$$X_n \Rightarrow X_\infty$$

as $n \to \infty$, where $X_\infty \stackrel{D}{=} 2(U - \frac{1}{2})$ and U is uniform on $[0,1]$. Note that if $X_0 = 1$, then the X_n's take values in the dyadic rationals D. Hence,

$$\frac{1}{n}\sum_{j=0}^{n-1} P(X_j \in D \,|\, X_0 = 1) \to 1$$

as $n \to \infty$, whereas $P(X_\infty \in D) = 0$ (because D is countable). This implies that

$$\frac{1}{n}\sum_{j=0}^{n-1} P(X_j \in \cdot \,|\, X_0 = 1)$$

does not converge in total variation norm to $\pi(\cdot) = P(X_\infty \in \cdot)$, which (in turn) implies that $\{X_n : n \geq 0\}$ cannot be shift-coupled (see Aldous and Thorisson [2]).

Our goal here is to (partially) address this issue by suggesting an alternative estimation strategy for such problems. We require that the state-space S of our process X be a metric space, equipped with metric ρ. (An important special case is that in which S is Euclidian with ρ equal to Euclidian norm.) Our key assumption will be the following:

A5. We can simulate a process $\{(X'(t), X^*(t)) : t \geq 0\}$ such that

1. $X'(t) \stackrel{D}{=} X(t)$ for $t \geq 0$;
2. $\{X^*(t) : t \geq 0\}$ is a stationary version of X under which $X(0)$ has distribution π; and
3. $\rho(X^*(t), X'(t)) \to 0$ a.s. as $t \to \infty$.

This assumption is closely related to the notion of ϵ-coupling (for details, see Asmussen [3]). Let

$$Y' = \int_{[0,\infty)} f(X'(s))G(ds),$$

$$Y^* = \int_{[0,\infty)} f(X^*(s))G(ds),$$

and consider the "controlled" estimator

$$Y_3(\lambda) = Y' - \lambda(Y^* - \nu g), \tag{6.18}$$

where $g = \int_{[0,\infty)} G(ds)$, and $\lambda \in \Re$. The idea is that if f is suitably continuous, then Assumption A5.3 will guarantee that Y' and Y^* will be close to one another and, hence, highly correlated. This is known to be a highly favorable setting for the use of control variates (see, e.g., Bratley, Fox, and Schrage [8]). By suitable continuity, we mean the following:

A6. $P(X^*(0) \in D_f) = 0$, where D_f is the set of discontinuities of $f: S \to \Re$.

If f is bounded, then Assumption A6, in conjunction with Assumption A5.3, guarantees that $Ef(X'(t)) - Ef(X^*(t)) \to 0$ as $t \to \infty$ (see Billingsley [7]).

The choice of λ in Eq. (6.1) can, of course, be optimized so as to minimize the variance of $Y_3(\lambda)$. Assume that var $Y' < \infty$, $0 <$ var $Y^* < \infty$. The optimal choice of λ is then

$$\lambda_* = \text{cov}(Y', Y^*)/\text{var } Y^*,$$

with a resulting minimizing variance of

$$\text{var } Y_3(\lambda_*) = \text{var } Y'(1 - \text{corr}(Y', Y^*)^2),$$

where $\text{corr}(Y', Y^*)^2$ is the squared coefficient of correlation between Y' and Y^* given by $\text{cov}(Y', Y^*)^2/\text{var } Y' \text{ var } Y^*$. Although λ_* is not known a priori, it can easily be estimated from a sample of i.i.d. replicates from the population (Y', Y^*). This estimation comes at no loss of asymptotic efficiency.

In particular, let $\alpha_3(c)$ be the estimator constructed in c units of computer time by independently replicating the random vector (Y', Y^*) and estimating λ_* in the obvious way (by replacing population moments by sample moments). Suppose $\tau_3(i)$ is the time required to generate the ith replicate $(Y_3'(i), Y_3^*(i))$. Then, if the appropriate moments are finite,

$$c^{1/2}(\alpha_3(c) - \alpha) \Rightarrow \sigma_3 N(0,1) \tag{6.19}$$

as $c \to \infty$, where $\sigma_3^2 = E\tau_3(1)\text{var } Y_3(\lambda_*)$. We can use Eq. (6.19) to analyze the (large t) efficiency of our proposed estimator $\alpha_3(c)$.

Example 1 (continued): First, observe that the computational effort $\tau_3(1,t)$, for this problem, grows linearly in t so that $E\tau_3(1,t) \sim c_3 t$ as $t \to \infty$ for some $c_3 > 0$. Furthermore, because λ_* is the variance minimizer, var $Y_3(\lambda_*) \leq$ var $Y_3(1)$, and

$$Y_3(1) = \nu g + f(X'(t)) - f(X^*(t)).$$

Because $Y_3(1) \to \nu g$ as $t \to \infty$, it follows (assuming uniform integrability) that var $Y_3(\lambda_*, t) \to 0$ as $t \to \infty$. We conclude that $\sigma_3^2 = \sigma_3^2(t) = o(t)$ as $t \to \infty$, so that $\alpha_3(c)$ is, for large c, a more efficient estimator than is $\alpha(c)$.

Example 2 (continued): Again, $E\tau_3(1,t) \sim c_3 t$ for some $c_3 > 0$. As in the analysis of Example 1, we take advantage of the fact that var $Y_3(\lambda_*) \leq$ var $Y_3(1)$, where

$$Y_3(1) = \nu g + \int_0^t [f(X'(s)) - f(X^*(s))]\, ds.$$

Suppose, as is often reasonable, that

$$f(X'(s)) - f(X^*(s)) = O(s^{-p}) \quad \text{for some } p > \tfrac{1}{2}.$$

Then, $\int_0^t [f(X'(s)) - f(X^*(s))]\, ds = o(t^{1/2})$ as $t \to \infty$. Hence, we can typically expect that var $Y_3(1) = o(t)$ as $t \to \infty$, from which it is evident that $\sigma_3^2(t) = o(t^2)$ as $t \to \infty$. This establishes that $\alpha_3(c)$ is more efficient than the naive estimator for large c.

We conclude this section with a brief discussion of how to generate (X', X^*) satisfying Assumption A5. We will restrict our attention to general state-space chains $\{X_n : n \geq 0\}$ satisfying stochastic recursions of the form

$$X_{n+1} = h(X_n, V_{n+1}),$$

where $h : S \times S' \to S$ and $\{V_n : n \geq 1\}$ is a sequence of i.i.d. S'-valued random elements (independent of X_0). For each $x \in S$, define $X_n(x)$ for $n \geq 0$ via $X_0(x) = x$ and

$$X_{n+1}(x) = h(X_n(x), V_{n+1})$$

for $n \geq 0$. A large class of such stochastic recursions have the property of the following:

A7. For each $x, y \in S, \rho(X_n(x), X_n(y)) \to 0$ a.s. as $n \to \infty$.

See, for example, Glynn [10]; the class of recurrent autoregressive sequences (either scalar or vector-valued) also have this property.

To construct (X', X^*), suppose that X_0' and X_0^* are generated independently, and set

$$X_{n+1}' = h(X_n', V_{n+1}),$$
$$X_{n+1}^* = h(X_n^*, V_{n+1}),$$

for $n \geq 0$; in other words, one drives both X' and X^* using the same sequence of V_i's. (This is an application of common random numbers.) Due to Assumption A7, it then follows that $\rho(X_n', X_n^*) \to 0$ a.s. as $n \to \infty$.

7. COMPUTATIONAL EXPERIENCE

For obvious reasons, it is desirable to perform our numerical experimentation on a model for which both the transient and steady-state distributions are analytically calculable. Consequently, we have chosen to focus our efforts on the M/M/∞ continuous-time birth–death process $X = (X(t) : t \geq 0)$ having birth rates $\lambda_n = \lambda$ and death rates $\mu_n = n\mu$ for $n \geq 0$. It is well known that if $X(0) = 0$ then $X(t)$ is Poisson distributed with parameter $(\lambda(1 - \exp(-\mu t))/\mu)$. In addition, we have chosen to use the performance measure $f(x) = x$, so that

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

TABLE 1. $\lambda = 0.5$, $\mu = 1.0$; Computer Budget $c = 100{,}000$

Time t	True Value	$\alpha(c)$	No. of Replications	95% CI	$\alpha_1(c)$	No. of Replications	95% CI
0.5	0.197	0.197	200,000	[0.195, 0.199]	0.196	201,604	[0.193, 0.198]
1.0	0.316	0.316	100,000	[0.313, 0.320]	0.318	170,547	[0.315, 0.320]
2.0	0.432	0.430	50,000	[0.424, 0.436]	0.434	147,660	[0.432, 0.435]
5.0	0.497	0.488	20,000	[0.479, 0.498]	0.497	138,422	[0.497, 0.497]
10.0	0.500	0.486	10,000	[0.472, 0.500]	0.500	137,867	[0.500, 0.500]

TABLE 2. $\lambda = 2.0$, $\mu = 1.0$; Computer Budget $c = 100{,}000$

Time t	True Value	$\alpha(c)$	No. of Replications	95% CI	$\alpha_1(c)$	No. of Replications	95% CI
0.5	0.787	0.789	200,000	[0.785, 0.793]	0.783	120,082	[0.775, 0.792]
1.0	1.264	1.261	100,000	[1.254, 1.268]	1.263	85,468	[1.254, 1.272]
2.0	1.729	1.730	50,000	[1.718, 1.742]	1.735	67,266	[1.728, 1.742]
5.0	1.987	1.984	20,000	[1.964, 2.003]	1.987	60,463	[1.986, 1.989]
10.0	2.000	2.009	10,000	[1.982, 2.037]	2.000	60,283	[2.000, 2.000]

Table 3. $\lambda = 0.5$, $\mu = 1.0$; Computer Budget $c = 100{,}000$

Time t	True Value	$\alpha(c)$	95% CI	No. of Replications	$\alpha_1(c)$	95% CI	No. of Replications	$\alpha_2(c)$	95% CI	No. of Replications
1	0.184	0.184	[0.182, 0.186]	100,000	0.184	[0.181, 0.186]	170,547	0.180	[0.176, 0.185]	63,650
5	0.401	0.395	[0.391, 0.400]	20,000	0.402	[0.400, 0.403]	138,422	0.400	[0.396, 0.404]	17,970
10	0.450	0.445	[0.439, 0.450]	10,000	0.450	[0.450, 0.451]	137,867	0.449	[0.447, 0.452]	9,457
20	0.475	0.472	[0.467, 0.478]	5,000	0.475	[0.475, 0.476]	137,847	0.477	[0.475, 0.478]	4,867

Table 4. $\lambda = 2.0$, $\mu = 1.0$; Computer Budget $c = 100{,}000$

Time t	True Value	$\alpha(c)$	95% CI	No. of Replications	$\alpha_1(c)$	95% CI	No. of Replications	$\alpha_2(c)$	95% CI	No. of Replications
10	1.800	1.799	[1.788, 1.810]	10,000	1.803	[1.800, 1.805]	60,283	1.805	[1.794, 1.817]	7,328
50	1.960	1.958	[1.946, 1.971]	2,000	1.961	[1.960, 1.961]	60,283	1.964	[1.959, 1.969]	1,863
100	1.980	1.981	[1.968, 1.993]	1,000	1.980	[1.980, 1.980]	60,283	1.980	[1.976, 1.983]	1,863
200	1.990	1.991	[1.980, 2.003]	500	1.990	[1.990, 1.990]	60,283	1.989	[1.987, 1.991]	491

To offer a fair comparison of coupling-based estimators relative to the naive (conventional) estimator, we need to ensure that both estimators are constructed from computer budgets that are (at least) roughly equal. Because of the machine-dependent subtleties that arise in explicitly timing these routines, we prefer to make the simplifying assumption that the computational effort can be taken equal to the total time simulated. For example, to estimate $\alpha = Ef(X(t))$ via the naive algorithm involves using t units of computer budget per replication, whereas the coupling-based estimator $\alpha_1(c)$ (based on independent coupling) requires $2(T_i \wedge t)$ units of computer budget for the ith replication (where T_i is the coupling time associated with replication i).

Tables 1 and 2 are concerned with comparing the naive estimator to the coupling-based estimator (under the independent coupling) for the performance measure $\alpha = EX(t)$.

Note that the performance of the naive estimator, as measured by the confidence interval half-width, tends to degrade as $t \to \infty$ (due to the corresponding decrease in the sample size as t gets large). On the other hand, the sample size associated with the coupling-based estimator $\alpha_1(c)$ declines as $t \to \infty$, whereas the variance of $\alpha_1(c)$ has a tendency to decrease. Thus, no monotonic trend in half-width is evident for $\alpha_1(c)$ (although one is guaranteed that the half-width gets smaller for t sufficiently large).

We turn now to the shift-coupling estimator $\alpha_2(c)$. Let X and f be as earlier, and consider the cumulative cost $\alpha = E\int_0^t f(X(s))\,ds$. Tables 3 and 4 display the results obtained for estimating α via the naive estimator $\alpha(c)$, the coupling-based estimator $\alpha_1(c)$, and the shift-coupling estimator $\alpha_2(c)$. The coupling-based estimator $\alpha_1(c)$ takes advantage of the independent coupling, whereas the shift-coupling estimator $\alpha_2(c)$ simulates X' via $X'(s) = X^*(T^* + s)$ for $s \geq 0$ (as suggested in Section 5).

Qualitatively, the results mimic those obtained for the performance measure $\alpha = EX(t)$. In particular, both the coupling-based estimator $\alpha_1(c)$ and the shift-coupling estimator $\alpha_2(c)$ have efficiencies that appear to improve as the time horizon t increases, just as our theory predicts. Furthermore, the efficiency of $\alpha_1(c)$ increases relative to $\alpha_2(c)$, in accordance with the theory of Section 5. (It is worth noting that in Table 4 all the replications coupled before $t = 10$; this explains why the number of replications computed for $c = 100{,}000$ is identical for the time horizons listed.)

References

1. Aldous, D. & Diaconis, P. (1987). Strong uniform times and finite random walks. *Advances in Applied Mathematics* 8: 69–97.
2. Aldous, D. & Thorisson, H. (1993). Shift-coupling. *Stochastic Processes and Their Applications* 1: 1–14.
3. Asmussen, S. (1992). Coupling and weak convergence. *Annals of Applied Probability* 2: 739–751.
4. Asmussen, S., Glynn, P.W., & Thorisson, H. (1992). Stationary detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation* 3: 130–157.

5. Athreya, K.B. & Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society* 245: 493–501.
6. Baskett, F., Chandy, K.M., Muntz, R.R., & Palacios, F. (1975). Open, closed, and mixed networks of queues with different classes of customers. *Journal of the Association for Computing Machines* 22: 248–260.
7. Billingsley, P. (1968). *Convergence of probability measures.* New York: John Wiley.
8. Bratley, P., Fox., B.L., & Schrage, L. (1987). *A guide to simulation.* New York: Springer-Verlag.
9. Fox, B.L. & Glynn, P.W. (1989). Estimating discounted costs. *Management Science* 35: 1297–1325.
10. Glynn, J.E. (1989). A discrete-time storage process with a general release rule. *Journal of Applied Probability* 26: 566–583.
11. Glynn, P.W. (1994). Some topics in regenerative steady-state simulation. *Acta Applicandae Mathematica* 34: 225–236.
12. Glynn, P.W. & L'Ecuyer, P. (1994). Likelihood ratio gradient estimation for stochastic recursions. *Advances in Applied Probability* (to appear).
13. Glynn, P.W. & Whitt, W. (1992). The asymptotic efficiency of simulation estimators. *Operations Research* 40: 505–520.
14. Kalashnikov, V. (1994). *Topics on regenerative processes.* Boca Raton, FL: CRC Press.
15. Kelly, F.P. (1979). *Reversibility and stochastic networks.* New York: John Wiley.
16. Kelly, F.P. (1986). Blocking probabilities in large circuit-switched networks. *Advances in Applied Probability* 18: 473–505.
17. Kielson, J. (1979). Markov Chain models—Rarity and exponentiality. New York: Springer-Verlag.
18. Lindvall, T. (1992). *Lectures on the coupling method.* New York: John Wiley.
19. Meyn, S.P. & Tweedie, R.L. (1993). *Markov Chains and stochastic stability.* New York: Springer-Verlag.
20. Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwunde Gebiete* 43: 309–318.
21. Sigman, K. (1990). One-dependent regenerative processes and queues in continuous time. *Mathematics of Operations Research* 15: 175–189.
22. Whittle, P. (1986). *Systems in stochastic equilibrium.* New York: John Wiley.

Probability in the Engineering and Informational Sciences, **10**, 1996, 187–196. Printed in the U.S.A.

AN IMPROVEMENT TO THE TOTAL HAZARD METHOD FOR SYSTEM RELIABILITY SIMULATION

HÉCTOR CANCELA
IRISA, Campus Universitaire de Beaulieu
35042 Rennes CEDEX, France
and
Inv. Operativa-InCo-PeDeCiBa
Facultad de Ingeniería
CP 11300 Montevideo, Uruguay

MOHAMED EL KHADIRI
IRISA, Campus Universitaire de Beaulieu
35042 Rennes CEDEX, France
and
Département Gestion Logistique et Transport
Institut Universitaire de Technologie
58 rue Michel Ange
44606 Saint-Nazaire, France

The total hazard Monte Carlo estimator has been successfully applied to estimate the unreliability of multicomponent systems. Nevertheless, there are some cases where it is less efficient than the crude estimator. We propose in this paper a simple modification that improves the performance of the total hazard estimator.

1. INTRODUCTION

Consider a coherent binary stochastic system S with m components such that each of them is either up or down (see, e.g., Barlow and Proschan [2]). The

This research was supported by the ECOS French-Uruguayan Scientific Cooperation Program (Action U93E03) and BID/CONICYT project 153 (Uruguay).

 0269-9648/96 $11.00 + .10