Proceedings of the 1992 Winter Simulation Conference
ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson

CORRELATION OF MARKOV CHAINS SIMULATED IN PARALLEL

Paul Glasserman

Columbia Business School
403 Uris Hall
New York, NY 10027, U.S.A.

Pirooz Vakili

Department of Manufacturing Engineering
Boston University
Boston, MA 02215, U.S.A.

ABSTRACT

We investigate the dependence induced among multiple Markov chains when they are simulated in parallel using a shared Poisson stream of potential event occurrences. One expects this dependence to facilitate comparisons among systems; our results support this intuition. We give conditions on the transition structure of the individual chains implying that the coupled process is an *associated* Markov chain. Association implies that variance is reduced in comparing increasing functions of the chains, relative to independent simulations, through a routine argument. We also give an apparently new application of association to the problem of selecting the better of two systems from limited data. Under conditions, the probability of incorrect selection is asymptotically smaller when the systems compared are associated than when they are independent. This suggests a further advantage to linking multiple systems through parallel simulation.

1 INTRODUCTION

Most work on parallel simulation stresses efficiency in evaluating the performance of a single system. The implications of parallelism for the comparison of several systems have received less attention (exceptions include Heidelberger and Nicol 1991, Ho, Sreenivas, and Vakili 1992, and Vakili 1992). When multiple systems are simulated together, in parallel, their outputs often become dependent, and this dependence must be considered in the statistical evaluation of comparisons.

Our purpose here is to examine the dependence introduced among multiple Markov chains when they are coupled through parallel simulation. One expects this dependence to facilitate comparisons; we give conditions that validate this intuition. Our model of parallelism is rather simple: we assume that each chain is uniformizable and that the various chains share a single Poisson stream of potential event times; this is the method of Vakili (1992). It seems reasonable to expect our results to extend, at least qualitatively, to other implementations based on uniformization, such as those described in Heidelberger and Nicol (1991).

Our results are based on conditions ensuring that the coupled process obtained by simulating multiple chains in parallel is an *associated* Markov chain. Association is a strong form of positive dependence, implying that all increasing functions of the various chains are positively correlated. Our use of association in this setting is similar to that in the analysis of *common random numbers* of Heidelberger and Iglehart (1979) and Glasserman and Yao (1992), but the conditions used here are quite different from the ones in those papers. A key difference is that we establish association by directly examining the transition structure of the coupled chain, rather than by attempting to show that the coupled process is an increasing function of i.i.d. random variables. A principal contribution of this paper is to identify conditions on the individual chains ensuring that the coupled chain is associated.

It is easy to show that a class of comparisons are statistically more efficient when the different chains are associated than when they are independent. This is one sense in which the coupling induced by parallel simulation is advantageous. We outline another sense. Suppose the goal is to select the system with the best performance. Under reasonable assumptions, the probability of failing to select the best system from finite simulation runs vanishes exponentially as the number of runs grows. For pairwise comparisons, we argue that when the systems compared are associated, the exponential rate is at least as great as when they are independent.

Section 2 formalizes our model of parallel simulation. Section 3 reviews association and its connection

with monotone Markov chains. Section 4 puts conditions on the transition structure of individual chains ensuring that the coupled chain is associated. Section 5 looks at the implications of association for correctly selecting the better of two systems.

2 MARKOV CHAINS SIMULATED IN PARALLEL

In this section, we describe a mechanism for simulating M Markov chains in parallel. The basic idea is to use a single shared clock that drives all chains simultaneously. The justification for this construction of chains is the well-known uniformization procedure. We begin with a brief review of this procedure.

2.1 Uniformization

Let $X = \{X_t, t \geq 0\}$ be a continuous-time Markov chain (CTMC) on a (finite or) countable set S. Denote by Q_{ij} the rate of transitions from state s_i to state s_j, and let $q_i = -Q_{ii}$ be the total rate of transitions out of s_i. We assume that Q is bounded (i.e., uniformizable), meaning that $\sup q_i < \infty$. Given that the chain is in some state, say s_i, it remains in that state for a duration that is exponentially distributed with mean q_i^{-1}. These sojourn times of the chain in states can be *uniformized* by appropriately introducing extra fictitious transitions from states to themselves. The inter-event times can thus be made an i.i.d. sequence of exponential random variables independent of the states of the chain.

More precisely, let $X = \{X_t, t \geq 0\}$ be a CTMC with infinitesimal generator Q bounded by Λ. Let $N^\Lambda = \{N_t, t \geq 0\}$ be a Poisson process with rate Λ and $Y = \{Y_n, n \geq 0\}$ a discrete-time Markov chain (DTMC) with transition probability matrix $P = I + \Lambda^{-1}Q$ (where I is the identity), with N^Λ and Y mutually independent. If X_0 has the distribution of Y_0, then $\{X_t, t \geq 0\}$ and $\{Y_{N_t}, t \geq 0\}$ are equal in law. The Poisson process N^Λ determines the (potential) state transition epochs of the CTMC X, while the state transitions are determined by the DTMC Y. Multiple Markov chains simulated in parallel can use the same dominating Poisson process.

2.2 Simulating One Markov Chain

Before discussing simulation of multiple Markov chains, we describe a specific implementation of uniformization available for many physically meaningful models. Our central assumption is that all state transitions can be classified into finitely many types of *events*, with the current state and the event type completely determining the next state. (So far, the only restriction this imposes is that there be an upper bound on the number of transitions out of any state. Later, we put more detailed conditions on the transitions.) This additional structure is often present in simulated systems, but is suppressed by the matrix Q.

As before, let S denote the state space. Let $E = \{e^1, \ldots, e^K\}$ be the set of events, with K finite. To each event e there corresponds a deterministic state transition function

$$f_e : S \to S.$$

Notice that f_e is defined on all of S. If event e is not active in state s, then $f_e(s) = s$, corresponding to a null transition.

For each event e^i, let λ_i be the maximum possible rate of event e^i and N^{λ_i} a Poisson process with rate λ_i, dominating the instances of e^i. These Poisson processes are assumed to be independent. The superposition of these processes is a Poisson process N^Λ with rate $\Lambda = \sum_{i=1}^K \lambda_i$, and the original processes can be recovered by thinning N^Λ (with probability λ_i/Λ to recover N^{λ_i}). This leads to the following model of the simulation clock.

Let $(\tau, \varepsilon) = \{(\tau_n, \varepsilon_n), n \geq 0\}$ be a marked Poisson process, where $\{\tau_n, n \geq 0\}$ is the sequence of arrival instances of N^Λ, and $\{\varepsilon_n, n \geq 0\}$ is an i.i.d. sequence of discrete random variables, independent of the Poisson process N^Λ, such that $\varepsilon_n \in E$ and $P(\varepsilon_n = e^i) = \lambda_i/\Lambda$. Λ is the rate at which the clock ticks, τ_n is the n-th tick of the clock, and ε_n is the type of event that occurs at the n-th tick of the clock.

Given initial state $Y_0 = X_0$, the state of the system evolves as follows:

$$Y_n = f_{\varepsilon_n} \circ f_{\varepsilon_{n-1}} \circ \cdots \circ f_{\varepsilon_1}(X_0)$$

and

$$X_t = \sum_{n=0}^{\infty} Y_n I\{\tau_n \leq t < \tau_{n+1}\}, \text{ for } t \geq 0$$

2.3 Simulating Multiple Markov Chains

We now use this mechanism to simulate M Markov chains simultaneously. Let

$$\begin{aligned} S^j &= \text{the state space of system } j; \\ E^j = E &= \text{event set of system } j; \\ f_e^j &= \text{state transition rule for event } e \text{ in system } j. \end{aligned}$$

By possibly enlarging some event sets, we can always assume (as above) that all M systems have the same event sets.

Define $F_e : \prod_1^M S^j \to \prod_1^M S^j$ componentwise:

$$F_e(x^1, \ldots, x^M) = (f_e^1(x^1), \ldots, f_e^M(x^M))$$

Given $X_0 = (X_0^1, \ldots, X_0^M)$, define $X = \{X_t, t > 0\}$ on $S \subseteq \prod_1^M S^j$ by

$$Y_n = F_{\varepsilon_n} \circ F_{\varepsilon_{n-1}} \circ \ldots \circ F_{\varepsilon_1}(X_0),$$

and

$$X_t = \sum_{n=0}^{\infty} Y_n I\{\tau_n \leq t < \tau_{n+1}\}, \text{ for } t \geq 0,$$

This construction defines a coupling of the M chains with each component process X^j having the correct marginal probability law. In general, the state space S of $\{X_t, t \geq 0\}$ is a strict subset of $\prod_1^M S^j$.

A note on implementation is in order. The above construction can be easily implemented in a variety of computational environments. In this model of simulation, at each tick of the clock, the present time (i.e., τ_n) and the type of the event (i.e., ε_n) are announced to all systems. In a serial implementation on a single processor, $F_e(x^1, \ldots, x^M)$ is implemented sequentially in a loop that executes $f_e^j(x^j)$ for $j = 1, \ldots, M$. In a parallel implementation, $F_e(x^1, \ldots, x^M)$ is implemented in a distributed fashion where each $f_e^j(x^j)$ (or possibly a group of them) is executed at a separate processor simultaneously and in parallel. In a SIMD (single-instruction, multiple-data) implementation, the clock is implemented at the front-end computer and each $f_e^j(x^j)$ is executed at a processor of the SIMD machine. At each tick of the clock, the time and type of event are broadcast to all processors. The processors then execute their respective $f_e^j(x^j)$ in a SIMD fashion. For further discussion on the computational aspects of this approach see Vakili (1992).

We now turn our attention to another potential advantage of this approach and to the main question we address in this paper. Since the M chains are simulated simultaneously and in parallel, it is possible to compare their performance simultaneously and in parallel. Does coupling the chains facilitate their comparison, compared to, say, simulating them independently?

Let L^i be a sample statistic from a finite-horizon simulation of system i. Then

$$\mathsf{Var}[L^i - L^j] = \mathsf{Var}[L^i] + \mathsf{Var}[L^j] - 2\mathsf{Cov}[L^i, L^j]. \quad (1)$$

To the extent that the coupling introduces positive covariance among the sample statistics, it reduces variance in (pairwise) comparisons, relative to independent simulations.

Equation (1) motivates an examination of when (and in what sense) Markov chains simulated with a shared clock are positively dependent. In particular, we develop conditions for these chains to exhibit *association*, a strong type of positive dependence. Association implies variance reduction in the setting of (1) and related comparisons. In Section 5, we outline another dividend of association.

3 ASSOCIATION AND MARKOV CHAINS

We now review some basic properties of association and conditions for a Markov chain to be associated. Association was introduced in Esary et al. (1967) as a property of sets of (real-valued) random variables: they defined the random variables $\{X_1, \ldots, X_n\}$ to be associated if all increasing functions of these variables are positively correlated; i.e., if

$$\mathsf{Cov}[f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)] \geq 0$$

for all increasing f and g for which the covariance exists. Esary et al. summarize simple properties of associated random variables. Among these are the following: subsets of associated random variables are associated; independent random variables are associated; increasing functions of associated random variables are associated; and a set consisting of a single random variable is associated. Association has proved to be a useful condition in many settings, including reliability, interacting particle systems, and the analysis of variance reduction techniques.

The utility of association is enhanced through a connection with a class of Markov chains. Daley (1968) defined a Markov chain on $\mathbf{R}$ to be *monotone* if its transition kernel P satisfies

$$x \leq y \Rightarrow P(x, [z, \infty)) \leq P(y, [z, \infty)), \forall z \in \mathbf{R}. \quad (2)$$

He noted that a monotone Markov chain $\{X_n, n \geq 0\}$ is an associated sequence, in the sense that all finite subsets of $\{X_n, n \geq 0\}$ are associated. The condition in (2) could alternatively be written as $P(X_1 \geq z | X_0 = x) \leq P(X_1 \geq z | X_0 = y)$ whenever $x \leq y$. An equivalent characterization is that $\mathsf{E}[f(X_1)|X_0 = x]$ is an increasing function of x for all bounded, increasing functions f. In some ways the most natural characterization is this: a Markov chain on $\mathbf{R}$ is monotone if for any pair $x, y \in \mathbf{R}$ with $x \leq y$ it is possible to construct two copies of the chain, $\{X_n^x, n \geq 0\}$ and $\{X_n^y, n \geq 0\}$ with $X_0^x = x$ and $X_0^y = y$, such that $X_n^x \leq X_n^y$ for all n. Just such a construction, starting from an i.i.d. sequence of uniform random variables, is carried out in Heidelberger and Iglehart (1979) as part of their analysis

of common random numbers. Since this construction transforms independent random variables monotonically to $\{X_n, n \geq 0\}$ it actually proves that a monotone Markov chain is an associated sequence.

Analogous properties and definitions apply in continuous time. A Markov process $\{X_t, t \geq 0\}$ on $\mathbf{R}$ is monotone if for all $0 \leq t_1 < t_2$ the transition kernel $P_{t_1 t_2}$ given by $P_{t_1 t_2}(x, A) = P(X_{t_2} \in A | X_{t_1} = x)$ is monotone in the sense of (2). This condition admits a sample-path interpretation just like the one given above for discrete-time chains. Any finite subset of a monotone Markov process is associated, so in this sense monotone Markov processes are associated processes. For Markov processes on finite sets, Keilson and Kester (1977) give conditions on the infinitesimal generator for monotonicity.

If we simulate M Markov processes $\{X_t^i, t \geq 0\}$, $i = 1, \ldots, M$, in parallel, the resulting coupled process $\{(X_t^1, \ldots, X_t^M), t \geq 0\}$ will often be a Markov process as well. In particular, this holds using the set-up of Section 2.3. Even if each X^i is a scalar process, the coupled process is vector-valued; so, we need conditions for association in higher dimensions.

We restrict attention to subsets of $\mathbf{R}^d$ though most properties we discuss apply to more general partially ordered sets. We assume $\mathbf{R}^d$ is endowed with a partial order $\leq$, which need not be the usual componentwise order, though it often will be. A set A is called an *upper set* (with respect to $\leq$) if $x \in A$ and $x \leq y$ together imply $y \in A$; thus, an upper set cannot be exited through upward movement. A probability measure μ on $\mathbf{R}^d$ is called associated if

$$\mu(A_1 \cap A_2) \geq \mu(A_1)\mu(A_2) \tag{3}$$

for all upper sets A_1, A_2 (Lindqvist 1988). This says that indicator functions for upper sets are positively correlated under μ. An equivalent characterization is that all bounded, increasing functions are positively correlated under μ. A random vector on $\mathbf{R}^d$ is called associated if its distribution is associated. Perhaps the most important distinction between association on $\mathbf{R}$ and association on $\mathbf{R}^d$, $d \geq 2$, is that in higher dimensions the set consisting of a single random element may not be associated, as it is in $\mathbf{R}$. To put it more generally, on totally ordered state spaces all probability measures are associated, but not so on partially ordered spaces.

Condition (2) has a natural generalization to $\mathbf{R}^d$. A transition kernel P and the corresponding chain $\{X_n, n \geq 0\}$ are called monotone if

$$x \leq y \Rightarrow P(x, A) \leq P(y, A), \quad \text{for all upper sets } A.$$

It is still true that $\mathsf{E}[f(X_1)|X_0 = x]$ increases in x for all increasing f, and the sample-path characterization remains valid as well. However, on a partially ordered set a monotone Markov chain may not be associated — a stronger condition is needed. If P is monotone and if, in addition, every probability measure $P(x, \cdot)$, $x \in \mathbf{R}^d$ is associated in the sense of (3), then the chain is associated as well as monotone. This is proved in Lindqvist (1988).

As before, conditions for continuous-time processes can be reduced to discrete time by considering transition operators from one fixed time to another. For our purposes, it is more convenient to restrict attention to countable partially ordered sets and work with conditions on the infinitesimal generator. We only consider bounded generators; i.e., those that are uniformizable. Following Massey (1987), we say that a generator Q is monotone if, for all upper sets A containing either both x and y or neither,

$$x \leq y \Rightarrow Q(x, A) \leq Q(y, A). \tag{4}$$

Monotone generators indeed generate monotone Markov processes. A generator is called *up-down* if it permits transitions only between comparable states; i.e., if $Q(x, y) > 0$ implies that either $x \leq y$ or $y \leq x$. The key result linking monotone Markov processes on countable sets and association is this:

Lemma 3.1. A Markov process $\{X_t, t \geq 0\}$ on a countable partially ordered set S with bounded generator is associated for all associated initial distributions if and only if its generator is monotone and up-down.

The original version of this result, requiring a finite state space, was given in Harris (1977); a far more general result, not even requiring a countable state space, appears in Liggett (1985, p.80).

4 ASSOCIATION OF PARALLEL MARKOV CHAINS

Let $X^1, \ldots, X^M$ be M Markov chains as in Section 2, X^j having countable state space S^j. Each S^j is assumed to be partially ordered. To avoid cumbersome notation, we use $\leq$ to denote the partial order on all state spaces. These partial orders are not necessarily identical (nor are the S^j's). Let $X = (X^1, \ldots, X^M)$ be the coupled process on $S \subseteq \prod_{j=1}^M S^j$ as defined in Section 2. We assume that the partial order on S is the *componentwise* order which we also denote by $\leq$.

We derive our result on the association of the coupled chain by focusing on the transition kernels. We begin with the following lemma that establishes a useful relation between properties of the generator of a continuous-time chain and the transition kernel of the corresponding uniformized discrete-time chain:

Lemma 4.1. Assume that a generator Q and a transition kernel P are related via $P = I + \Lambda^{-1}Q$, I the identity matrix. Then
(i) If P is up-down, Q is up-down.
(ii) If P is monotone, Q is monotone.

Proof. From $Q(x,y) > 0$ it follows that $x \neq y$. Note that for $x \neq y$, $P(x,y) = \Lambda Q(x,y)$ and $Q(x,y) > 0$ implies $P(x,y) > 0$. Since P is up-down, $P(x,y) > 0$ yields $x \geq y$ or $x \leq y$. Therefore $Q(x,y) > 0$ yields $x \geq y$ or $x \leq y$ and hence Q is up-down. This proves (i).

Assume that $x \leq y$, and A is an upper set that excludes both x and y. Then

$$P(x,A) = \sum_{z \in A} P(x,z) = \sum_{z \in A} I(x,z) + \frac{1}{\Lambda} \sum_{z \in A} Q(x,z) = \frac{1}{\Lambda} Q(x,A).$$

Similarly

$$P(y,A) = \frac{1}{\Lambda} Q(y,A)$$

hence, if P is monotone,

$$P(x,A) \leq P(y,A) \Rightarrow Q(x,A) \leq Q(y,A)$$

If A includes both x and y then

$$P(x,A) = \sum_{z \in A} P(x,z) = \sum_{z \in A} I(x,z) + \frac{1}{\Lambda} \sum_{z \in A} Q(x,z) = 1 + \frac{1}{\Lambda} Q(x,A).$$

Similarly

$$P(y,A) = 1 + \frac{1}{\Lambda} Q(y,A)$$

therefore

$$P(x,A) \leq P(y,A) \Rightarrow Q(x,A) \leq Q(y,A)$$

hence Q is a monotone generator and (ii) is proved. □

Consider a single Markov chain defined via the representation in Section 2. Define f_e to be up-down if $f_e(x) \leq x$ or $f_e(x) \geq x$ for all $x \in S$.

Lemma 4.2. If f_e is increasing and up-down for all $e \in E$, then the transition kernel P for the discrete-time process $\{Y_n, n \geq 0\}$ is up-down and monotone.

Proof. For $x, z \in S$ define $P_e(x,z)$ by

$$P_e(x,z) = \begin{cases} 1 & \text{if } f_e(x) = z \\ 0 & \text{if } f_e(x) \neq z \end{cases}$$

and let $p_i = P(\varepsilon_n = e^i)$. Then the transition kernel P is determined via

$$P(x,z) = \sum_{i=1}^{K} p_i P_{e^i}(x,z)$$

Thus if $P(x,z) > 0$ then $P_e(x,z) > 0$ for some $e \in E$; hence $z = f_e(x)$ for some $e \in E$ and since f_e is up-down $z \leq x$ or $z \geq x$. Therefore, P is up-down.

Now assume $x \leq y$; then by hypothesis $f_e(x) \leq f_e(y)$ for all $e \in E$. For any upper set A, $f_e(x) \in A$ implies that $f_e(y) \in A$ by the definition of upper sets. Note that $P_e(x,A) = 1$ if $f_e(x) \in A$ and $P_e(x,A) = 0$ if $f_e(x) \notin A$. Hence, the conclusion in the previous sentence can be re-written as: $P_e(x,A) = 1$ implies $P_e(y,A) = 1$. Therefore, we have

$$x \leq y \Rightarrow P_e(x,A) \leq P_e(y,A), \quad \text{for any } e \in E.$$

Note that $P(x,A) = \sum_{i=1}^{K} p_i P_{e^i}(x,A)$. Hence, if $x \leq y$,

$$\begin{aligned} P(x,A) &= \sum_{i=1}^{K} p_i P_{e^i}(x,A) \\ &\leq \sum_{i=1}^{K} p_i P_{e^i}(y,A) = P(y,A). \end{aligned}$$

Therefore, P is a monotone kernel. □

Combining Lemmas 4.1 and 4.2 we obtain the following result:

Theorem 4.3. If f_e is increasing and up-down for all $e \in E$, then the generator Q of the continuous-time process $X = \{X_t, t \geq 0\}$ is up-down and monotone, hence the process X is associated for all associated initial distributions.

Now we turn to the construction of M chains in parallel. To ensure that the coupled process $X = (X^1, \ldots X^M)$ is associated we make F_e up-down and increasing. Keeping in mind that S is typically a strict subset of $\prod_1^M S^j$, let us say that f_e^j, $j = 1, \ldots, M$ are up-down *in the same direction* on S if for all $x = (x^1, \ldots, x^M) \in S$, either

$$f_e^j(x^j) \geq x^j, \quad j = 1, \ldots, M,$$

or

$$f_e^j(x^j) \leq x^j, \quad j = 1, \ldots, M.$$

The following theorem is the main result of this section:

Theorem 4.4. If f_e^j are increasing functions for all $e \in E$ and $1 \leq j \leq M$, and if for each $e \in E$,

the functions f_e^j, $j = 1\ldots, M$, are up-down in the same direction on S, then the coupled process $X = (X^1, \ldots, X^M)$ is associated for all associated initial distributions.

Proof. Since the partial order on S is the componentwise order, it is trivial to verify that f_e^j increasing for $1 \le j \le M$ implies that F_e is increasing. Also if f_e^js are "up" ("down") for $1 \le j \le M$ then F_e is "up" ("down"). Hence by Lemma 4.2 X is associated. □

We now point out an easy consequence of association. Fix a (deterministic) time-horizon $T > 0$. Suppose L^i is an increasing real-valued function of $\{X_t^i, 0 \le t \le T\}$, interpreted as a cost or performance measure for system i. Then association of $\{X_t, t \ge 0\}$ implies that L^i and L^j are positively correlated, for all i and j. Via (1), we get

Corollary 4.5. If L^i, $i = 1, \ldots, M$ is an increasing function of $\{X_t^i, 0 \le t \le T\}$, then under the conditions of Theorem 4.4, $\mathsf{Var}[L^i - L^j] \le \mathsf{Var}[L^i] + \mathsf{Var}[L^j]$. Thus, coupling the chains reduces variance compared to independent simulation.

Under appropriate additional conditions, it is possible to extend Corollary 4.5 to steady-state estimates and to multiple comparisons.

5 PROBABILITY OF CORRECT SELECTION

The previous section showed that parallel simulation of Markov processes resulting in an associated coupled process reduces variance in making comparisons. We now consider a different setting and establish another consequence of association. Suppose that from the processes we simulate in parallel we wish to choose the one with best performance; e.g., the one maximizing some expectation. With finite simulation runs, there is typically some probability that the process with the best sample performance is not the one with the best expected performance. In this case, picking the best observed system may result in an incorrect selection. We show for pairwise comparisons that if the systems compared are associated, then the probability of incorrect selection is asymptotically smaller than if the simulations are carried out independently.

To develop this idea in more detail, we restrict attention to the case of two discrete-time processes $X = \{X_n, n \ge 0\}$ and $Y = \{Y_n, n \ge 0\}$. We allow X and Y to be dependent and denote by $\tilde{X}$ and $\tilde{Y}$ a pair of independent copies of X and Y. Suppose that

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f(X_i) = \mu_X$$

and

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f(Y_i) = \mu_Y$$

exist almost surely and are constants. Suppose that $\mu_X > \mu_Y$ and that this makes system X preferable to system Y. After simulating the two systems for a finite time, we pick whichever has the greatest observed performance. The events

$$G_n = \{n^{-1} \sum_{i=1}^{n} f(X_i) - n^{-1} \sum_{i=1}^{n} f(Y_i) < 0\}$$

and

$$\tilde{G}_n = \{n^{-1} \sum_{i=1}^{n} f(\tilde{X}_i) - n^{-1} \sum_{i=1}^{n} f(\tilde{Y}_i) < 0\}$$

are the events of incorrect selection using, respectively, coupled and independent simulations. Various decision rules have been proposed and extensively analyzed in the simulation literature, some with the explicit goal of maximizing the probability of correct selection; for a recent survey, see Goldsman et al. (1991). A different perspective on related issues is given in Ho et al. (1992).

Ideally, we would like conditions on the dependence of X and Y under which $P(G_n) \le P(\tilde{G}_n)$. Intuitively, positive dependence between the two processes would seem to support this inequality, but association by itself does not appear to imply it, in general. However, both $P(G_n)$ and $P(\tilde{G}_n)$ vanish as n increases; in fact, under mild conditions both go to zero exponentially fast. We argue that, with association, the exponential rate for $P(G_n)$ is at least as great as that for $P(\tilde{G}_n)$.

We consider only the case of an i.i.d. sequence $\{(X_n, Y_n), n \ge 0\}$; at the end of this section we comment on the Markov case. Define the moment generating functions

$$M_X(\theta) = \mathsf{E}[e^{\theta f(X_0)}], \quad M_Y(\theta) = \mathsf{E}[e^{\theta f(Y_0)}],$$

$$M(\theta) = \mathsf{E}[e^{\theta[f(X_0) - f(Y_0)]}],$$

and

$$\tilde{M}(\theta) = \mathsf{E}[e^{\theta[f(\tilde{X}_0) - f(\tilde{Y}_0)]}] = M_X(\theta) M_Y(-\theta).$$

We assume that M_X and M_Y are finite for every θ in a neighborhood of the origin. This implies that M and $\tilde{M}$ are also finite in some neighborhood of the origin. Association enters through the following simple observation:

Lemma 5.1. If (X_0, Y_0) is an associated vector and if f is an increasing function, then $M(\theta) \le \tilde{M}(\theta)$.

Proof. For all $\theta \in \mathbf{R}$, one of the two functions $x \mapsto \exp(\theta x)$ and $x \mapsto \exp(-\theta x)$ is increasing and the other is decreasing. If (X_0, Y_0) is associated and f is increasing, then $f(X_0)$ and $f(Y_0)$ are associated. But then $\exp(\theta f(X_0))$ and $\exp(-\theta f(Y_0))$ are negatively correlated; i.e.,

$$\mathsf{E}[e^{\theta[f(X_0)-f(Y_0)]}] - \mathsf{E}[e^{\theta f(X_0)}]\mathsf{E}[e^{-\theta f(Y_0)}] \leq 0,$$

as claimed. □

Define events F_n and $\tilde{F}_n$ by replacing the strict inequalities in G_n and $\tilde{G}_n$ with weak inequalities. We now have

Theorem 5.2. There exist non-negative constants c and $\tilde{c}$ such that the following hold:

$$\begin{aligned}
\limsup_{n\to\infty} n^{-1} \log P(F_n) &\leq -c \\
\liminf_{n\to\infty} n^{-1} \log P(G_n) &\geq -c \\
\limsup_{n\to\infty} n^{-1} \log P(\tilde{F}_n) &\leq -\tilde{c} \\
\liminf_{n\to\infty} n^{-1} \log P(\tilde{G}_n) &\geq -\tilde{c}.
\end{aligned}$$

The constants satisfy $\tilde{c} \leq c$.

Proof. Define

$$I(x) = \sup_{\theta}\{\theta x - \log M(\theta)\},$$

and define $\tilde{I}$ from $\tilde{M}$ in the same way. It follows from Chernoff (1952, Theorem 1) that the inequalities in the statement of the theorem hold with

$$c = \inf_{x<0} I(x) = \inf_{x\leq 0} I(x)$$

and

$$\tilde{c} = \inf_{x<0} \tilde{I}(x) = \inf_{x\leq 0} \tilde{I}(x).$$

So, it suffices to show that $I(x) \geq \tilde{I}(x)$ for all x. But this follows from the fact that $M(\theta) \leq \tilde{M}(\theta)$ for all θ. □

The limits in the theorem essentially state that, using associated samples,

$$P(\text{error}) \sim e^{-nc},$$

and using independent samples

$$P(\text{error}) \sim e^{-n\tilde{c}},$$

Since $c \geq \tilde{c}$, the probability of incorrect selection is asymptotically smaller with association.

There are extensions of Chernoff's theorem for sequences more general than i.i.d., including the Markov case. Verification of the conditions needed to generalize Theorem 5.2 is the subject of current investigation.

6 SUMMARY

We have investigated the dependence induced among multiple Markov chains simulated in parallel using a shared stream of potential event times. Conditions on the transition structure of the individual chains guarantee that the coupled process is associated. Association leads to variance reduction in comparing the performance of the various systems. It also suggests that the probability of failing to select the best system is asymptotically smaller (as the sample size grows) when the chains are coupled than when they are simulated independently.

ACKNOWLEDGMENT

P.G. is supported, in part, by the National Science Foundation under grant ECS-9216490

REFERENCES

Chernoff, H., A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations, *Ann. Math. Stat.* **23**, 493-507, 1952.

Daley, D., Stochastically Monotone Markov Chains, *Z. Wahrsch. verw. Geb.*, **10**, 305-317, 1968.

Esary, J.D., Proschan, F., and Walkup, D.W., Association of Random Variables, *Ann. Math. Stat.*, **38**, 1466-1474, 1967.

Glasserman, P., and Yao, D.D., Some Guidelines and Guarantees for Common Random Numbers, *Mgmt. Sci.* **38**, 884-908, 1992.

Goldsman, D., Nelson, B., and Schmeiser, B., Methods for Selecting the Best System, *Proceedings of the 1991 Winter Simulation Conference,* 177-186, The Society for Computer Simulation, San Diego, CA, 1991.

Harris, T.E., A Correlation Inequality for Markov Processes on Partially Ordered Spaces, *Ann. Prob.* **5**, 451-454, 1977.

Heidelberger, P., and Iglehart, D.L., Comparing Stochastic Systems Using Regenerative Simulation and Common Random Numbers, *Adv. Appl. Prob.* **11**, 804-819, 1979.

Heidelberger, P., and Nicol, D.M., Simultaneous Parallel Simulations of Continuous Time Markov Chains at Multiple Parameter Settings, *Proceedings of the 1991 Winter Simulation Conference,* 602-607, The Society for Computer Simulation, San Diego, CA, 1991.

Ho, Y.C., Sreenivas, R., and Vakili, P., Ordinal Optimization of Discrete Event Dynamic Systems, *DEDSTA*, to appear, 1992.

Keilson, J., and Kester, A., Monotone Matrices and Monotone Markov Processes, *Stoch. Proc. Appl.* **5**, 231-241, 1977.

Liggett, T.M., *Interacting Particle Systems,* Springer, New York, 1985.

Lindqvist, B.H., Association of Probability Measures on Partially Ordered Sets, *J. Multivariate Anal.,* **26**, 111-132, 1988.

Massey, W.A., Stochastic Orderings for Markov Processes on Partially Ordered Spaces, *Math. Oper. Res.,* **12**, 350-367, 1987.

Vakili, P., Massively Parallel and Distributed Simulation of a Class of Discrete-Event Systems: a Different Perspective, TOMACS, to appear, 1992.

AUTHOR BIOGRAPHIES

PAUL GLASSERMAN is an Associate Professor in the Management Science division of Columbia Business School. In 1988-1990, he was a Member of Technical Staff in the Operations Research department of AT&T Bell Laboratories. His simulation-related research interests include variance reduction and gradient estimation.

PIROOZ VAKILI is an Assistant Professor in the department of Manufacturing Engineering at Boston University. His research interests include gradient estimation, stochastic optimization, and parallel and distributed simulation.