

Ariel

**Rapid #: -3996253****IP: 128.59.152.191****1**

Status	Rapid Code	Branch Name	Start Date
Pending	AUM	FC Depository	12/17/2010 8:49:18 AM

**CALL #:****UA030001 per****LOCATION:****AUM :: FC Depository :: FIVE COLLEGE  
DEPOSITORY**

TYPE: Article CC:CCL  
JOURNAL TITLE: Probability in the engineering and informational sciences  
USER JOURNAL TITLE: Probability in the Engineering and Informational Sciences  
AUM CATALOG TITLE: Probability in the engineering and informational sciences.  
ARTICLE TITLE: Comparing Markov chains simulated in parallel.  
ARTICLE AUTHOR: Glasserman, Paul, and Pirooz Vakili  
VOLUME: 8  
ISSUE: 3  
MONTH:  
YEAR: 1994  
PAGES: 309-326  
ISSN: 0269-9648  
OCLC #: AUM OCLC #: 15354091  
CROSS REFERENCE ID: 256754  
VERIFIED:

**BORROWER:****ZCU :: Butler  
Wiehardt, Ginny****PATRON:**

PATRON ID:  
PATRON ADDRESS:  
PATRON PHONE:  
PATRON FAX:  
PATRON E-MAIL: glw2113@columbia.edu  
PATRON DEPT: Libraries  
PATRON STATUS: Officer  
PATRON NOTES:



This material may be protected by copyright law (Title 17 U.S. Code)  
System Date/Time: 12/17/2010 12:51:08 PM MST

Editor  
**Sheldon Ross**  
University of California, Berkeley

Editorial Board

**M. Brown**  
The City College, CUNY, U.S.A.

**F. K. Hwang**  
Bell Laboratories, U.S.A.

**F. Kelly**  
Cambridge University, U.K.

**F. J. Radermacher**  
Universität Ulm, Germany

**G. Shanthikumar**  
University of California, Berkeley,  
U.S.A.

**K. Ross**  
University of Pennsylvania, U.S.A.

**H. M. Taylor**  
University of Delaware, U.S.A.

**H. C. Tijms**  
The Free University, The Netherlands

**S. Xu**  
Pennsylvania State University, U.S.A.

**R. Weber**  
Cambridge University, U.K.

**G. Weiss**  
Georgia Tech University, U.S.A.

**D. Yao**  
Columbia University, U.S.A.

**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).

**1994 Subscription Information:** *Probability in the Engineering and Informational Sciences* (ISSN 0269-9648) is published quarterly. Institutional rate: US \$231.00 in the U.S.A. and Canada; UK £160.00 in all other countries. Individual rates (prepayment by personal check or credit card only): US \$69.00 in the U.S.A. and Canada, UK £54.00 in all other countries. Prices include postage and insurance; air mail UK £14.50 per year extra.

**Reprints** are available from the authors. Orders of 50 or more reprints, contact the Journals Production Office.

Copyright © 1994 Cambridge University Press

All rights reserved. No part of this publication may be reproduced, in any form or by any means, electronic, photocopying or otherwise, without permission in writing from the copyright holder.

**Photocopying information for users in the USA:** The Item-Fee Code for this publication indicates that copying for internal or personal use beyond that permitted by Sec. 107 or 108 of the US Copyright Law is authorized for users duly registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the appropriate remittance of \$5.00 per article is paid directly to: CCC, 222 Rosewood Drive, Danvers, MA 01923. Specific written permission must be obtained for all other copying; Contact your nearest Cambridge University Press office. *The Item-Fee Code* for this journal is: 0269-9648/94 \$5.00 + .00

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.

## COMPARING MARKOV CHAINS SIMULATED IN PARALLEL

PAUL GLASSERMAN\*

Columbia Business School, 403 Uris Hall, New York, New York 10027

PIROOZ VAKILI\*\*

Manufacturing Engineering Department, Boston University,  
Boston, Massachusetts 02215

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

\*Research supported by NSF grant MSS-9216490.

\*\*Research supported by NSF grants DDM-9215368 and EID-9212122.

of several systems have received less attention; exceptions include Heidelberger and Nicol [9], Ho, Sreenivas, and Vakili [10], and Vakili [19]. 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 [19]. 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 [9].

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 [8] and Glasserman and Yao [5], 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 introduce 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.

A preliminary version of this paper appeared in the *Proceedings of the Winter Simulation Conference* (Glasserman and Vakili [4]). This paper extends the earlier one in several respects, most notably in Section 5.

## 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 transition from state  $s_i$  to state  $s_j$ , and let  $q_i = -Q_{ii}$  be the total rate of transition 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  $\Lambda$ . Let  $N^\Lambda = \{N_t, t \geq 0\}$  be a Poisson process with rate  $\Lambda$  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  $N^\Lambda$  and  $Y$  mutually independent. If  $X_0$  has the distribution of  $Y_0$ , then  $\{X_t, t \geq 0\}$  and  $\{Y_n, n \geq 0\}$  are equal in law. The Poisson process  $N^\Lambda$  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 representation of the uniformized chain that corresponds to the simulation implementation discussed in this paper.

Our central assumption is that all state transitions can be classified into finitely many types of *events* where the current state and the event type determine the next state, in general probabilistically. The only restriction this imposes is that there be an upper bound on the number of transitions out of any state. This additional structure is often present in many physically meaningful models but is suppressed by the matrix  $Q$ . It is worth mentioning that the following representation is not unique. Several representations of the same chain can be defined, some more appropriate than others for specific purposes.

As before, let  $S$  denote the state space. Let  $E = \{e^1, \dots, e^K\}$  be the set of events, with  $K$  finite. To model probabilistic transitions explicitly, we assume

that to each event  $e$  and each  $u \in [0,1]$ , there corresponds a state transition function

$$f_{(e,u)} : S \rightarrow S,$$

with the interpretation that event  $e$  and random number  $u$  change the system state from  $s$  to  $f_{(e,u)}(s)$ . Notice that  $f_{(e,u)}$  is defined on all of  $S$ . If event  $e$  is not active in state  $s$ , then  $f_{(e,u)}(s) = s$ , corresponding to a null transition.

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

Let  $(\tau, (\varepsilon, u)) = \{(\tau_n, (\varepsilon_n, u_n)), n \geq 0\}$  be a marked Poisson process, where  $\{\tau_n, n \geq 0\}$  is the sequence of arrival instances of  $N^\Lambda$ , and  $\{(\varepsilon_n, u_n), n \geq 0\}$  is an i.i.d. sequence of random vectors, independent of the Poisson process  $N^\Lambda$ , such that  $\varepsilon_n \in E$ ,  $P(\varepsilon_n = e^i) = \lambda_i/\Lambda$ , and  $u_n$  is uniformly distributed on  $[0,1]$ .  $\Lambda$  is the rate at which the clock ticks,  $\tau_n$  is the  $n$ th tick of the clock,  $\varepsilon_n$  is the type of event that occurs at the  $n$ th tick of the clock, and  $u_n$  is the random variate used to generate the  $n$ th state transition.

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

$$Y_n = f_{(\varepsilon_n, u_n)} \circ f_{(\varepsilon_{n-1}, u_{n-1})} \circ \cdots \circ f_{(\varepsilon_1, u_1)}(X_0)$$

and

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

To illustrate we give some examples. We return to these in subsequent sections.

**Example 1:** Consider a tandem queueing network consisting of  $K$  exponential servers, the  $i$ th having rate  $\mu_i$ . Assume a Poisson stream of arrivals (with rate  $\lambda$ ) to the network. Let  $B_i$  be the buffer space (including the space at the server) preceding server  $i+1$  ( $B_i \leq \infty$ ,  $i = 1, \dots, K$ ). Assume service at a server does not begin if the immediate downstream buffer is full (the so-called communication blocking). In this case, we may take  $S = \{(x_1, \dots, x_K); 0 \leq x_i \leq B_i\}$ ,  $E = \{a, d_i, i = 1, \dots, K\}$ , where  $a$  = arrival to the network and  $d_i$  = departure from server  $i$ .

**Example 2:** Consider a reliability system consisting of  $K$  binary components, i.e., the state space of the system is  $S = \{0,1\}^K$ . Assume the lifetime of each component is exponentially distributed (with rate  $\lambda_i$  for component  $i$ ) and that there are  $m$  repair facilities for repairing failed components (using some scheduling scheme for the order in which failed components are repaired). Assume

the repair times of the components are exponentially distributed, with rate  $\mu_j$  for resource  $j$ . Let  $X_t = (x_1(t), \dots, x_K(t))$  be the state of the system at time  $t$ . Note that the processes  $\{x_i(t); t \geq 0\}$ ,  $i = 1, \dots, K$ , may be dependent due to, e.g., a limited number of repair facilities or the policy used to repair components. We assume that the process  $X = \{X_t; t \geq 0\}$  is Markov. In this case, let  $E = \{e_i, r_j, i = 1, \dots, K, k = 1, \dots, m\}$ , with  $e_i$  = failure of component  $i$  and  $r_k$  = end of repair of a component by repair facility  $k$ . The rates of events are as already given.

### 2.3. Simulating Multiple Markov Chains

We now use the preceding construction to simulate  $M$  Markov chains simultaneously. Let

$S^j$  = the state space of system  $j$ ,

$E^j = E$  = event set of system  $j$ , and

$f_{(e,u)}^j$  = state transition rule for event  $e$  in system  $j$ .

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

Define  $F_{(e,u)} : \prod_1^M S^j \rightarrow \prod_1^M S^j$  componentwise:

$$F_{(e,u)}(x^1, \dots, x^M) = (f_{(e,u)}^1(x^1), \dots, f_{(e,u)}^M(x^M)).$$

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

$$Y_n = F_{(\varepsilon_n, u_n)} \circ F_{(\varepsilon_{n-1}, u_{n-1})} \circ \cdots \circ F_{(\varepsilon_1, u_1)}(X_0)$$

and

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

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 preceding construction can easily be implemented in a variety of computational environments. In this model of simulation, at each tick of the clock, the present time (i.e.,  $\tau_n$ ), the type of the event (i.e.,  $\varepsilon_n$ ), and the uniform variate (i.e.,  $u_n$ ) are announced to all systems. In a serial implementation on a single processor,  $F_{(e,u)}^j(x^1, \dots, x^M)$  is implemented sequentially in a loop that executes  $f_{(e,u)}^j(x^j)$  for  $j = 1, \dots, M$ . In a parallel implementation,  $F_{(e,u)}(x^1, \dots, x^M)$  is implemented in a distributed fashion where each  $f_{(e,u)}^j(x^j)$  (or possibly a group of them) is executed at a separate processor simultaneously and in parallel. In a single-instruction, multiple-data (SIMD) implementation, the clock is implemented at the front-end computer and each  $f_{(e,u)}^j(x^j)$  is executed at a processor of the SIMD

machine. At each tick of the clock, the time, type of event, and the uniform variate are broadcast to all processors. The processors then execute their respective  $f_{(e,u)}^j(x^j)$  in a SIMD fashion. For further discussion on the computational aspects of this approach, see Vakili [19].

*Example 1 (continued):* Multiple versions of the tandem queueing network of Example 1 can be considered that differ in their service rates, number of buffers, or arrival rate. Let  $\lambda^j$  be the rate of arrival and  $\mu_i^j$  the service rate of the  $i$ th server at the  $j$ th version of the network. The rates of these events at the shared clock are, respectively,  $\lambda = \max\{\lambda^j; j = 1, \dots, M\}$  and  $\mu_i = \max\{\mu_i^j; j = 1, \dots, M\}$ .

*Example 2 (continued):* Consider different versions of the reliability system of Example 2 where the differences between versions are, e.g., in rates of failures of components, repair rates, number of components or repair facilities, or the scheduling policy for repairing components. If  $\lambda_i^j$  and  $\mu_k^j$  are the rates of events  $e_i$  and  $r_k$  at version  $j$ , then the rates of these events at the shared clock are, respectively,  $\lambda = \max\{\lambda_i^j; j = 1, \dots, M\}$  and  $\mu_k = \max\{\mu_k^j; j = 1, \dots, M\}$ .

We now turn our attention to another potential advantage of this approach and to the main question we address in this paper. Because 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 simulation of system  $i$ . Then

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

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 Eq. (1) and related comparisons. In Section 5 we investigate 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, Proschan, and Walkup [2] as a property of sets of (real-valued) random variables: they defined the random variables  $\{X_1, \dots, X_n\}$  to be associated if all increasing functions of these variables are positively correlated, i.e., if

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

for all increasing  $f$  and  $g$  for which the covariance exists. Esary et al. [2] summarized 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 [1] defined a Markov chain on  $\mathbf{R}$  to be *monotone* if its transition kernel  $P$  satisfies

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

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 Eq. (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  $E[f(X_1) | X_0 = x]$  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 [8] as part of their analysis of common random numbers. Because 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 Eq. (2). This condition admits a sample-path interpretation just like that already given 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 [13] gave conditions on the infinitesimal generator for monotonicity.

If we simulate  $M$  Markov processes  $\{X_t^i, t \geq 0\}$ ,  $i = 1, \dots, M$ , in parallel, the resulting coupled process  $\{(X_t^1, \dots, X_t^M), t \geq 0\}$  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  $\mathbf{R}^d$ , although 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, although 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  $\mu$  on  $\mathbf{R}^d$  is called associated if

$$\mu(A_1 \cap A_2) \geq \mu(A_1)\mu(A_2) \quad (3)$$

for all upper sets  $A_1, A_2$  [15]. This says that indicator functions for upper sets are positively correlated under  $\mu$ . An equivalent characterization is that all bounded, increasing functions are positively correlated under  $\mu$ . 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.

Equation (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

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

It is still true that  $E[f(X_1) | X_0 = x]$  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 Eq. (3), then the chain is associated as well as monotone. This is proved in Lindqvist [15].

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 [16], we say that a generator  $Q$  is monotone if, for all upper sets  $A$  containing both  $x$  and  $y$  or neither,

$$x \leq y \Leftrightarrow Q(x, A) \leq Q(y, A). \quad (4)$$

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 the following lemma.

**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 [7]; a far more general result, not even requiring a countable state space, appears in Liggett [14, p. 80].

#### 4. ASSOCIATION OF PARALLEL MARKOV CHAINS

Let  $X^1, \dots, X^M$ , 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  $X = (X^1, \dots, X^M)$  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  $\leq$ .

##### 4.1. Conditions for Association

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 condition (i).

Assume that  $x \leq y$ ,  $A$  is an upper set, and  $x, y \in A$  or  $x, y \notin A$ . Clearly,  $I(x, A) = I(y, A)$ . Hence,  $P(x, A) \leq P(y, A)$  yields

$$Q(x, A) = \Lambda(P(x, A) - I(x, A)) \leq \Lambda(P(y, A) - I(y, A)) = Q(y, A);$$

therefore,  $P$  monotone ( $x \leq y \Leftrightarrow P(x, A) \leq P(y, A)$ ) implies  $Q$  monotone, and condition (ii) is proved. ■

Consider a single Markov chain defined via the representation in Section 2. Define  $f_e = f_{(e, \cdot)}$  to be up-down if either  $f_{(e, u)}(x) \leq x$  for all  $x \in S$  and all  $u \in [0, 1]$  or  $f_{(e, u)}(x) \geq x$  for all  $x \in S$  and all  $u \in [0, 1]$ . Define  $f_e$  to be increasing if  $x \leq y$  implies  $f_{(e, u)}(x) \leq f_{(e, u)}(y)$  for all  $u \in [0, 1]$ .

**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  $P_{(e, u)}(x, z)$  by

$$P_{(e, u)}(x, z) = \begin{cases} 1 & \text{if } f_{(e, u)}(x) = z \\ 0 & \text{if } f_{(e, u)}(x) \neq z \end{cases}$$

and  $P_e(x, z)$  by

$$P_e(x, z) = \int_0^1 P_{(e, u)}(x, z) du.$$

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

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

If  $P(x, z) > 0$ , then  $P_e(x, z) > 0$  for some  $e \in E$ ; hence,  $z = f_{(e, u)}(x)$  for some  $u \in [0, 1]$  and because  $f_e$  is up-down  $z \leq x$  or  $z \geq x$ . Therefore,  $P$  is up-down.

Now assume  $x \leq y$ ; then, by hypothesis  $f_{(e, u)}(x) \leq f_{(e, u)}(y)$  for all  $e \in E$  and  $u \in [0, 1]$ . For any upper set  $A$ ,  $f_{(e, u)}(x) \in A$  implies that  $f_{(e, u)}(y) \in A$  by the definition of upper sets. Note that  $P_{(e, u)}(x, A) = 1$  if  $f_{(e, u)}(x) \in A$  and  $P_{(e, u)}(x, A) = 0$  if  $f_{(e, u)}(x) \notin A$ . Hence, the conclusion in the previous sentence can be rewritten as  $x \leq y$  implies  $P_{(e, u)}(x, A) \leq P_{(e, u)}(y, A)$ . Therefore, we have

$$x \leq y \Rightarrow P_e(x, A) = \int_0^1 P_{(e, u)}(x, A) du \leq \int_0^1 P_{(e, u)}(y, A) du = P_e(y, A)$$

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

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

Therefore,  $P$  is monotone. ■

Combining Lemmas 4.1 and 4.2, we obtain the following theorem.

**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  $X = (X^1, \dots, X^M)$  is associated, we make  $F_e$  up-down and increasing. Keeping in mind that  $S$  is typically a strict subset of  $\prod_{j=1}^M S^j$ , let us say that  $f_e^j$ ,  $j = 1, \dots, M$ , are up-down in the same direction on  $S$  if for all  $x = (x^1, \dots, x^M) \in S$  and all  $u \in [0, 1]$  either

$$f_{(e, u)}^j(x^j) \geq x^j, \quad j = 1, \dots, M,$$

or

$$f_{(e, u)}^j(x^j) \leq x^j, \quad j = 1, \dots, M.$$

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 for each  $e \in E$ , the functions  $f_e^j$ ,  $j = 1, \dots, M$ , are up-down in the same*

direction on  $S$ , then the coupled process  $X = (X^1, \dots, X^M)$  is associated for all associated initial distributions.

**PROOF:** Because the partial order on  $S$  is the componentwise order, it is trivial to verify that  $f_e^j$  increasing for  $1 \leq j \leq M$  implies that  $F_e$  is increasing. The up-down property extends the same way from  $f^j$  to  $F$ . Hence, by Lemma 4.2  $X$  is associated. ■

**Example 1 (continued):** Consider  $M$  versions of the tandem queueing network introduced in Section 2.2. Let  $S^j = \{x^j = (x_1^j, \dots, x_K^j); 0 \leq x_i^j \leq B_i^j\}$  be the state space of the  $j$ th network. Define the following partial order on  $S^j$ :  $x^j \leq x'^j$  if and only if  $\{\sum_{i=r}^K x_i^j \leq \sum_{i=r}^K x_i'^j; r = 1, \dots, K\}$ . For tandem queueing networks this partial order is a useful one; e.g., in this setting the departure intensity of the network is an increasing function. With respect to this partial order all departure events cause a “down” move and the arrival event causes an “up” move. Moreover, it is simple to verify that the state transition functions associated with the events are all increasing. Using a shared clock, as in Section 2.3, any clock event, say  $\varepsilon_n$ , will be a departure event from the same server at all networks or an arrival event at all networks (in both cases the event may be a fictitious event at some networks). Hence, the transition functions corresponding to all events are up-down in the same direction on the state space of the coupled chain,  $S$ , and the conditions of Theorem 4.4 are satisfied.

**Example 2 (continued):** Consider  $M$  versions of the reliability system introduced in Section 2.2. Let  $S^j = \{0, 1\}^{K_j}$  be the state space of the  $j$ th system. Define the following partial order on  $S^j$ :  $x^j \leq x'^j$  if and only if  $x_i^j \leq x_i'^j$ ,  $i = 1, \dots, K_j$ . With respect to this partial order, all failure events cause a “down” move and all repair events cause an “up” move. Assuming that the component failures within each system occur independently of each other, it is easy to verify that for all failure events,  $e_i$ ,  $f_{e_i}^j$  is increasing. Assume that the repair scheduling policy at each system is such that the transition functions associated with the repairs are also increasing (this is not automatically guaranteed for all repair policies). Again, using a shared clock that synchronizes events across systems such that the same event occurs at all systems at the same time ensures that the transition functions corresponding to all events are up-down in the same direction on the state space of the coupled chain,  $S$ , and the conditions of Theorem 4.4 are satisfied.

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

**COROLLARY 4.5:** *If  $L^i$ ,  $i = 1, \dots, M$ , is an increasing function of  $\{X_t^i, t \geq 0\}$ , then under the conditions of Theorem 4.4,  $\text{Var}[L^i - L^j] \leq \text{Var}[L^i] + \text{Var}[L^j]$ . Thus, coupling the chains reduces variance compared to independent simulation.*

## 4.2. Discrete-Time Conversion

A common use of uniformization in the context of simulation of Markov chains is discrete-time conversion. We give a brief description and refer to Fox and Glynn [3] for details.

Let  $X = \{X_t; t \geq 0\}$  be a CTMC and let  $\{Y_n; n \geq 0\}$  be a corresponding uniformized discrete-time chain. Consider a finite deterministic horizon,  $[0, T]$ . Let  $N = N_T$  be the number of events during  $[0, T]$ . Let  $t_i = \tau_i - \tau_{i-1}$ ,  $i = 1, \dots, N$  and  $t_{N+1} = T - \tau_N$ . Let  $g: S \rightarrow \mathbf{R}$  be a real-valued function on the state space  $S$ , and let  $L$ , a performance index, be defined by

$$L = \int_0^T g(X_t) dt = \sum_{i=1}^{N+1} g(Y_{i-1}) t_i.$$

$N$  is a Poisson random variable with parameter  $\Lambda T$  and  $E[t_i | N] = T/(N+1)$ ,  $i = 1, \dots, N+1$ . Hence,

$$K = E[L | N, Y] = \frac{T}{N+1} \sum_{i=1}^{N+1} g(Y_{i-1}).$$

In the discrete-time conversion approach, samples of  $K$  are used to estimate  $\theta = E[L]$ . Under some conditions  $K$  is a preferred estimator of  $\theta$ , in comparison to  $L$ , because to obtain samples of  $K$  only the discrete-time chain  $\{Y_n; n \geq 0\}$  need be simulated and it can be further shown that  $\text{Var}(K) \leq \text{Var}(L)$  and, hence,  $K$  is a more accurate estimator of  $\theta$ .

The conditions of Theorem 4.4 do not guarantee association when  $M$  uniformized discrete-time chains  $\{Y_n^j; n \geq 0\}$ ,  $j = 1, \dots, M$ , are simulated in parallel. The up-down property and monotonicity of the transition kernel of a discrete-time chain are neither necessary nor sufficient for association of the chain (see Harris [7]).

In the proposition to follow, we show that the conditions of Theorem 4.4 lead to induction of positive correlation between the discrete-time statistics in some more restricted settings.

Let

$$L_r = \int_0^T g_r(X_t) dt = \sum_{i=1}^{N+1} g_r(Y_{i-1}) t_i, \quad r = 1, 2,$$

and let

$$K_r = E[L_r | N, Y], \quad r = 1, 2.$$

Assume that the functions  $g_r$  are bounded, i.e.,  $|g_r(x)| \leq C$ , for all  $x \in S$ , and  $r = 1, 2$ . Then we have the following proposition.

**PROPOSITION 4.6:** *If  $\text{Cov}(L_1, L_2) > 0$ , then  $\text{Cov}(K_1, K_2) > 0$  for all sufficiently large uniformization rates  $\Lambda$ .*

**PROOF:**

$$\begin{aligned} \text{Cov}(L_1, L_2) &= E[\text{Cov}(L_1, L_2 | N, Y)] + \text{Cov}(E[L_1 | N, Y], E[L_2 | N, Y]) \\ &= E[\text{Cov}(L_1, L_2 | N, Y)] + \text{Cov}(K_1, K_2) \end{aligned}$$

It can be shown that (see Karlin and Taylor [12, Ch. 13])

$$\text{Cov}(t_i, t_j | N) = -T^2 \frac{1}{(N+1)^2(N+2)} \quad \text{when } i \neq j$$

$$\text{Var}(t_i | N) = T^2 \frac{N}{(N+1)^2(N+2)}.$$

Hence,

$$\begin{aligned} \text{Cov}(L_1, L_2 | N, Y) &= \frac{T^2}{(N+1)^2(N+2)} \sum_{i=1}^{N+1} \sum_{j=1}^{N+1} (g_1(Y_{i-1})g_2(Y_{i-1}) - g_1(Y_{i-1})g_2(Y_{j-1})). \end{aligned}$$

Note that  $|g_1(Y_{i-1})g_2(Y_{i-1}) - g_1(Y_{i-1})g_2(Y_{j-1})| \leq 2C^2$ ; therefore,

$$\text{Cov}(L_1, L_2 | N, Y) \leq \frac{T^2}{(N+1)^2(N+2)} (2C^2(N+1)^2) = 2T^2C^2 \frac{1}{N+2}$$

and

$$E[\text{Cov}(L_1, L_2 | N, Y)] \leq BE \left[ \frac{1}{N+2} \right] = B \frac{e^{-\Lambda T} + \Lambda T - 1}{\Lambda^2},$$

for some constant  $B$ . From the preceding it can be deduced that

$$\text{Cov}(K_1, K_2) \geq \text{Cov}(L_1, L_2) - B \frac{e^{-\Lambda T} + \Lambda T - 1}{\Lambda^2}.$$

Hence, if  $\text{Cov}(L_1, L_2) > 0$ , there exists a sufficiently large  $\Lambda$  such that  $\text{Cov}(K_1, K_2) > 0$ . ■

## 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 type of comparison and establish a further 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 incorrect selection. We show 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 motivate an examination of asymptotic error probabilities, we first give an example to show that association does not suffice to reduce the probability of incorrect selection (relative to independent sampling) for all sample sizes. Let the random variables  $X_1, X_2$  take values in  $\{0, 1, 2\}$  according to the following joint distribution:

		$X_2$		
		0	1	2
$X_1$	0	0.2	0.2	0.1
	1	0.2	0	0
	2	0	0	0.3

A simple calculation verifies that  $X_1$  and  $X_2$  are associated and that  $X_2$  has the larger mean. Suppose we observe a sample of  $(X_1, X_2)$  and apply the following decision rule to the problem of determining which has the higher mean: choose the index of the larger value; in case of a tie, randomize. Then, the probability of correct selection is

$$P(X_1 < X_2) + \frac{1}{2}P(X_1 = X_2) = 0.55.$$

If, however,  $\tilde{X}_1$  and  $\tilde{X}_2$  are independent random variables with the marginal distributions of  $X_1$  and  $X_2$ , respectively, then the same rule applied to  $(\tilde{X}_1, \tilde{X}_2)$  results in a correct-selection probability of

$$P(\tilde{X}_1 < \tilde{X}_2) + \frac{1}{2}P(\tilde{X}_1 = \tilde{X}_2) = 0.56.$$

Thus, the positive dependence induced by association does not guarantee a lower error probability.

We will show, however, that association is advantageous in an asymptotic sense that seems well suited to simulation problems. To develop this idea in more detail while avoiding technical complications, we restrict attention to irreducible, finite-state Markov processes  $X = \{X_t, t \geq 0\}$  and  $\{Y_t, t \geq 0\}$ . We allow  $X$  and  $Y$  to be dependent and assume that the bivariate process  $(X, Y)$  is irreducible and Markov. We denote by  $\tilde{X}$  and  $\tilde{Y}$  a pair of independent copies of  $X$  and  $Y$ . Let  $f, g$  be increasing, real-valued functions on the state space of  $X$  and  $Y$ , respectively. Let

$$U_t \equiv t^{-1} \int_0^t f(X_u) du \rightarrow \mu$$

and

$$V_t \equiv t^{-1} \int_0^t g(Y_u) du \rightarrow \nu$$

almost surely, for constants  $\mu, \nu$ . Let  $\tilde{U}_t$  and  $\tilde{V}_t$  be the corresponding quantities for  $\tilde{X}$  and  $\tilde{Y}$ . Suppose that  $\mu > \nu$  leads us to prefer system  $X$  to system  $Y$ . After simulating the two systems for a finite time, we pick whichever has the greatest observed performance. The events

$$G_t = \{U_t - V_t < 0\}$$

and

$$\tilde{G}_t = \{\tilde{U}_t - \tilde{V}_t < 0\}$$

are the events of incorrect selection at time  $t$  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, Nelson, and Schmeiser [6]. A different perspective on related issues is given in Ho, Sreenivas, and Vakili [10].

Ideally, we would like conditions on the dependence of  $X$  and  $Y$  under which  $P(G_t) \leq P(\tilde{G}_t)$ . Intuitively, positive dependence between the two processes would seem to support this inequality, but the preceding example indicates that association by itself does not imply it. Both  $P(G_t)$  and  $P(\tilde{G}_t)$  vanish as  $n$  increases; in fact, both go to zero exponentially fast. We will show that, with association, the exponential rate for  $P(G_t)$  is at least as great as that for  $P(\tilde{G}_t)$ .

For the coupled process  $(X, Y)$ , define

$$P'_{XY}((x_1, y_1), (x_2, y_2), \theta) = E_{(x_1, y_1)}[e^{\theta(U_t - V_t)}; X_t = x_2, Y_t = y_2],$$

and for the marginal processes define

$$P'_X(x_1, x_2, \theta) = E_{x_1}[e^{\theta U_t}; X_t = x_2]$$

and

$$P'_Y(y_1, y_2, \theta) = E_{y_1}[e^{\theta V_t}; Y_t = y_2],$$

where in each case the subscript on the expectation operator indicates the initial state. Each of the matrices  $P'_{XY}(\theta)$ ,  $P'_X(\theta)$ , and  $P'_Y(\theta)$  is nonnegative and (when finite) has a maximal (Perron-Frobenius) eigenvalue. Let  $\lambda_{XY}(\theta)$ ,  $\lambda_X(\theta)$ , and  $\lambda_Y(\theta)$  denote these eigenvalues. Suppose that the domain of each (the set on which it is finite) is open.

These functions play a central role in the large deviations of Markov additive processes, through results of Miller [17], Iscoe, Ney, and Nummelin [11],

and Ney and Nummelin [18]. The results of Iscoe et al. are particularly suited to our setting. For real  $v$ , set

$$\psi^*(v) = \sup_{\theta} \{v\theta - \log \lambda_{XY}(\theta)\}$$

and

$$\tilde{\psi}^*(v) = \sup_{\theta} \{v\theta - \log \lambda_X(\theta) - \log \lambda_Y(-\theta)\}.$$

We now have the following theorem.

**THEOREM 5.1:** *With the preceding notation,*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log P(G_t) = -\psi^*(0) \quad (5)$$

and

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log P(\tilde{G}_t) = -\tilde{\psi}^*(0). \quad (6)$$

If  $(X, Y)$  is associated, then  $\psi^*(0) \geq \tilde{\psi}^*(0)$ .

The novelty of this result lies in the last statement, which follows from this.

**LEMMA 5.2:** *For all  $\theta$ ,  $\psi^*(\theta) \geq \tilde{\psi}^*(\theta)$ .*

**PROOF:** For all values of  $\theta$ , one of the two functions  $x \mapsto e^{\theta x}$  and  $y \mapsto e^{-\theta y}$  is increasing and the other is decreasing. Consequently, it follows from the association of  $(X, Y)$  and the monotonicity of  $f$  and  $g$  that

$$\mathbb{E}[e^{\theta(U_t - V_t)}] \leq \mathbb{E}[e^{\theta U_t}] \mathbb{E}[e^{-\theta V_t}]. \quad (7)$$

It follows from Iscoe et al. [11] that

$$\log \lambda_{XY}(\theta) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta(U_t - V_t)}],$$

$$\log \lambda_X(\theta) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta U_t}],$$

and

$$\log \lambda_Y(\theta) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta V_t}];$$

so we conclude from Eq. (7) that

$$\log \lambda_{XY}(\theta) \leq \log \lambda_X(\theta) + \log \lambda_Y(-\theta),$$

from which it follows that  $\psi^*(\theta) \geq \tilde{\psi}^*(\theta)$ . ■

We now turn to the final proof.

**PROOF OF THEOREM 5.1:** It only remains to verify that the limits in Eqs. (5) and (6) are as claimed. From Theorem 5.1 of Iscoe et al. [11], we know that

$$\liminf_{t \rightarrow \infty} \frac{1}{t} \log P(G_t) \geq -\inf_{v > 0} \psi^*(v)$$

and

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log P(F_t) \leq -\inf_{v \geq 0} \psi^*(v),$$

where  $F_t$  is defined like  $G_t$  but with a weak inequality in place of the strict inequality. In particular,  $P(F_t) \geq P(G_t)$ . By Theorem 5.2 of Iscoe et al. [11], zero is a *dominating point* for both these limits, so

$$\inf_{v > 0} \psi^*(v) = \inf_{v \geq 0} \psi^*(v) = \psi^*(0),$$

and Eq. (5) holds.

Because independence is a special case of association, Eq. (6) follows from Eq. (5). More precisely, if  $X$  and  $Y$  are independent, then  $\lambda_{XY}(\theta) = \lambda_X(\theta)\lambda_Y(-\theta)$ , so  $\psi^*$  reduces to  $\tilde{\psi}^*$  and Eq. (5) to Eq. (6). ■

## References

1. Daley, D. (1968). Stochastically monotone Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 10: 305-317.
2. Esary, J.D., Proschan, F., & Walkup, D.W. (1967). Association of random variables. *Annals of Mathematical Statistics* 38: 1466-1474.
3. Fox, B.L. & Glynn, P.W. (1990). Discrete-time conversion for simulating finite-horizon Markov processes. *SIAM Journal on Applied Mathematics* 50: 1457-1473.
4. Glasserman, P. & Vakili, P. (1992). Correlation of uniformized Markov chains simulated in parallel. *Proceedings of the Winter Simulation Conference*. San Diego, CA: Society for Computer Simulation, pp. 475-482.
5. Glasserman, P. & Yao, D.D. (1992). Some guidelines and guarantees for common random numbers. *Management Science* 38: 884-908.
6. Goldsman, D., Nelson, B., & Schmeiser, B. (1991). Methods for selecting the best system. *Proceedings of the 1991 Winter Simulation Conference*. San Diego, CA: Society for Computer Simulation, pp. 177-186.
7. Harris, T.E. (1977). A correlation inequality for Markov processes on partially ordered spaces. *Annals of Probability* 5: 451-454.
8. Heidelberger, P. & Iglehart, D.L. (1979). Comparing stochastic systems using regenerative simulation and common random numbers. *Advances in Applied Probability* 11: 804-819.
9. Heidelberger, P. & Nicol, D. (1991). Simultaneous parallel simulations of continuous time Markov chains at multiple parameter settings. *Proceedings of the 1991 Winter Simulation Conference*. San Diego, CA: Society for Computer Simulation, pp. 602-607.
10. Ho, Y.C., Sreenivas, R., & Vakili, P. (1992). Ordinal optimization of discrete event dynamic systems. *DESDTA* 2: 61-88.
11. Iscoe, I., Ney, P., & Nummelin, E. (1985). Large deviations of uniformly recurrent Markov additive processes. *Advances in Applied Mathematics* 6: 373-412.

12. Karlin, S. & Taylor, H.M. (1981). *A second course in stochastic processes*. New York: Academic Press.
13. Keilson, J. & Kester, A. (1977). Monotone matrices and monotone Markov processes. *Stochastic Processes and Their Applications* 5: 231–241.
14. Liggett, T.M. (1985). *Interacting particle systems*. New York: Springer.
15. Lindqvist, B.H. (1988). Association of probability measures on partially ordered sets. *Journal of Multivariate Analysis* 26: 111–132.
16. Massey, W.A. (1987). Stochastic orderings for Markov processes on partially ordered spaces. *Mathematics of Operations Research* 12: 350–367.
17. Miller, H.D. (1961). A convexity property in the theory of random variables on a finite Markov chain. *Annals of Mathematical Statistics* 32: 1260–1270.
18. Ney, P. & Nummelin, E. (1987). Markov additive processes, I. Eigenvalue properties and limit theorems, II. Large deviations. *Annals of Probability* 15: 561–609.
19. Vakili, P. (1992). Massively parallel and distributed simulation of a class of discrete event systems: A different perspective. *TOMACS* 2(3): 214–238.

## POLLING IN A CLOSED NETWORK

EITAN ALTMAN

INRIA  
Centre Sophia Antipolis  
06565 Valbonne Cedex, France

URI YECHIALI

Department of Statistics and Operations Research  
Tel-Aviv University  
Tel-Aviv 69978, Israel

We consider a closed queueing network with a fixed number of customers, where a single server moves cyclically between  $N$  stations, rendering service in each station according to some given discipline (Gated, Exhaustive, or the Globally Gated regime). When service of a customer (message) ends in station  $j$ , it is routed to station  $k$  with probability  $P_{jk}$ . We derive explicit expressions for the probability generating function and the moments of the number of customers at the various queues at polling instants and calculate the mean cycle duration and throughput for each service discipline. We then obtain the first moments of the queues' length at an arbitrary point in time. A few examples are given to illustrate the analysis. Finally, we address the problem of optimal dynamic control of the order of stations to be served.

### 1. INTRODUCTION

In polling systems that have been studied in the literature [e.g., 1,3,4,6–9], one usually asserts independent Poisson arrivals. However, there are situations where arrivals strongly depend on the departure process. One encounters such

Supported by a grant from the France–Israel Scientific Cooperation (in Computer Science and Engineering) between the French Ministry of Research and Technology and the Israeli Ministry of Science and Technology, grant 3321190.