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).

1995 Subscription Information: *Probability in the Engineering and Informational Sciences* (ISSN 0269-9648) is published quarterly. Institutional rate: US $250.00 in the U.S.A. and Canada; UK £169.00 in all other countries. Individual rates (prepayment by personal check or credit card only): US $73.00 in the U.S.A. and Canada, UK £56.00 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.

Probability in the Engineering and Informational Sciences, **9**, 1995, 497–523. Printed in the U.S.A.

ACCELERATED REGENERATION FOR MARKOV CHAIN SIMULATIONS

SIGRÚN ANDRADÓTTIR*†
Department of Industrial Engineering
University of Wisconsin–Madison
1513 University Avenue
Madison, Wisconsin 53706

JAMES M. CALVIN
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

PETER W. GLYNN‡
Department of Operations Research
Stanford University
Stanford, California 94305

This paper describes a generalization of the classical regenerative method of simulation output analysis. Instead of blocking a generated sample path on returns to a fixed return state, a more general scheme to randomly decompose the path is used. In some cases, this decomposition scheme results in regeneration times that are a supersequence of the classical regeneration times. This "accelerated" regeneration is advantageous in several simulation contexts. It is shown that when this decomposition scheme accelerates regeneration relative to the classical regenerative method, it also yields a smaller asymptotic variance of the regenerative variance estimator than the classical method. Several other contexts in which increased regeneration frequency is beneficial are also discussed.

*Supported by the National Science Foundation under grant DDM-9210679.

†Current address: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332.

‡Supported by the National Science Foundation under grant DDM-9101580 and the Army Research Office under contract DAAL03-91-G-0319.

0269-9648/95 $11.00 + .10

1. INTRODUCTION

The regenerative method of simulation output analysis exploits the fact that regenerative stochastic processes have sample paths that can be broken up into independent and identically distributed cycles, called regenerative cycles. Thus, the steady-state simulation problem can be reduced to the finite-horizon simulation problem of estimating expected values of functions of the stochastic process over a regenerative cycle. This is typically achieved by selecting a distinguished "return state" and then starting a new regenerative cycle every time the stochastic process of interest visits the return state. The purposes of this study are to consider other ways of defining regenerative cycles and to compare the quality of the resulting point and variance estimators used to construct confidence intervals.

Consider, for example, a finite state space irreducible Markov chain. The classical regenerative method chooses one state as the return state, and the regenerative cycle boundaries are then defined as the times at which the chain visits the return state. If there are N states, then there are N choices of return state, each having different statistical qualities in general. A difficulty in implementing the classical regenerative method is that when the number of states N is large there may not be a good choice of a return state, in that all states may have a low steady-state probability and are thereby not visited very often, so the regenerative cycles become very long. There are situations, however, where the standard way of defining the regenerative cycles results in very long cycles, even though the rate of convergence to steady state is fast. This is the case when the Markov chain exhibits a significant amount of "row similarity" (e.g., when the Markov chain describes a sequence of independent and identically distributed random variables). Methods in this paper address this issue by taking advantage of row similarity in certain Markov chains to obtain "fast regenerations." The generalized regenerative structure can be exploited in several important ways:

- It is possible to obtain regenerative variance estimators with lower asymptotic variance than classical regenerative estimators in many cases (see Theorem 1).
- In several simulation contexts, including parallel regenerative simulation (Section 3) and some terminating simulations (Section 8), there is significant practical advantage to reducing the regeneration cycle time.
- We show that there is a connection between the best possible tail behavior of the regenerative cycles constructed using our approach and the rate of convergence of the chain to stationarity (Theorem 3).

Section 2 introduces the basic idea and preliminary results. Section 3 compares the statistical qualities of standard and accelerated regenerative estimators. The ideas are extended to the general state space setting and to higher order minorizations in Section 4. In Section 5 the tail behavior of accelerated regeneration times is compared with the time to reach stationarity. Section 6 considers the extension to continuous time. Examples are presented in Section 7, and Section 8 discusses applications of the accelerated regeneration idea to transient simulation. All proofs are deferred to the Appendix.

2. ACCELERATED REGENERATION FOR FINITE STATE MARKOV CHAINS

Let $X = (X_n : n \geq 0)$ be an irreducible Markov chain with finite state space S. Let P and π denote the transition probability matrix and stationary distribution of the Markov chain X, respectively. The classical approach to constructing regenerative cycles for such a chain is to select a state $z \in S$ and to use the "return times"

$$T_0(z) = \inf\{m \geq 0 : X_m = z\}, \qquad T_n(z) = \inf\{m > T_{n-1}(z) : X_m = z\}$$

(for $n \geq 1$) to define the cycle boundaries. In particular, it is well known that the cycles $\mathcal{C}_i = (X_{T_{i-1}(z)}, \ldots, X_{T_i(z)-1})$ are independent and identically distributed random elements of $\mathcal{S} = \bigcup_{j=1}^{\infty} S^j$ for $i \geq 1$.

We will now describe a new class of regeneration times, based on randomized stopping times, that offer the potential to increase the rate at which X regenerates, thereby providing "accelerated regeneration." The following concept plays a key role in our analysis.

DEFINITION 1: *Suppose that* $0 \neq \lambda \triangleq (\lambda_x : x \in S) \geq 0$ *and that* φ *is a probability distribution on* S. *We call* (λ, φ) *a* minorization *of* P *if*

$$P_{xy} \geq \lambda_x \varphi_y \tag{1}$$

for all $x, y \in S$.

Relation (1) allows us to rewrite P_{xy} in the form

$$P_{xy} = \lambda_x \varphi_y + (1 - \lambda_x) Q_{xy} \tag{2}$$

for all $x, y \in S$, where $Q_{xy} \triangleq (P_{xy} - \lambda_x \varphi_y)/(1 - \lambda_x)$ if $\lambda_x < 1$ and $Q_{xy} \triangleq \varphi_y$ otherwise. By summing Eq. (1) over y, we conclude that $\lambda_x \leq 1$, and hence Eq. (2) expresses $(P_{xy} : y \in S)$ as a convex combination of the two probability distributions φ and $(Q_{xy} : y \in S)$.

Equation (2) suggests the following approach to simulating X. If the chain currently occupies $x \in S$, we generate a Bernoulli random variable with parameter λ_x. If the Bernoulli random variable takes on the value 1, we distribute X on the next transition according to φ. Otherwise, if the Bernoulli random variable takes on the value 0, we distribute X on the next transition according to $(Q_{xy} : y \in S)$. The times at which the chain is distributed according to φ form regeneration times for X, because the distribution φ of the chain at such times is independent of the chain's position at the previous step.

In contrast to the classical cycles described earlier, we note that the construction of this new class of cycles requires that one must not only simulate X

but also the sequence of Bernoulli random variables. Mathematically, this corresponds to the fact that classical regeneration times are stopping times with respect to the filtration $\mathcal{F}_n \triangleq \sigma(X_k : 0 \le k \le n)$, $n \ge 0$, whereas the new class of regenerations are, in general, only randomized stopping times with respect to $(\mathcal{F}_n : n \ge 0)$.

It should also be pointed out that this approach to constructing regenerations, while new in discrete state space, has long played an important role in the theory of Markov chains taking values in a general state space. Specifically, representation (2) and the ensuing construction of regeneration times goes back to Athreya and Ney [3] and Nummelin [20].

We are interested in studying the rate at which X regenerates under this new class of randomized regeneration times. If $K(n)$ denotes the number of such regenerations in the interval $[0, n]$, it is easy to show that

$$\frac{K(n)}{n} \to \sum_{x \in S} \pi_x \lambda_x \quad \text{a.s.} \tag{3}$$

as $n \to \infty$, and hence $\sum_{x \in S} \pi_x \lambda_x$ can be interpreted as the "rate" at which these regenerations occur. On the other hand, it is clear that the corresponding rate at which classical regenerations occur, based on returns to state z, is just π_z. To show that the use of the randomized stopping times discussed already "accelerates" regeneration, we must establish that there exist minorizations for which $\sum_{x \in S} \pi_x \lambda_x \ge \pi_z$.

We first note that Eq. (3) implies that we would ideally like to find a minorization (λ, φ) with λ as large as possible. Because Eq. (1) implies that

$$\lambda_x \le P_{xy}/\varphi_y$$

for all $y \in S$ (with $0/0 \triangleq +\infty$), it follows (for a given φ) that the maximal λ is given by

$$\lambda_x^* = \min_{y \in S} \frac{P_{xy}}{\varphi_y} \tag{4}$$

for all $x \in S$. From Eq. (3) it is clear that λ^* maximizes the rate at which regeneration occurs (for a given φ). In Section 3, we show that λ^* also minimizes the asymptotic variance of the regenerative variance estimator (again for a fixed φ).

From Eqs. (3) and (4), it is clear that the optimal minorization can be computed as the solution to the nonlinear programming problem:

$$\max_{\varphi} \sum_{x \in S} \pi_x \min_{y \in S} \frac{P_{xy}}{\varphi_y} \tag{5}$$

$$\text{subject to:} \quad \sum_{y \in S} \varphi_y = 1, \quad \text{and} \quad \varphi_y \ge 0, \quad \text{for all } y \in S.$$

In Section 4 we prove that this problem always has at least one solution (see Theorem 2). We shall have more to say about this optimization problem in Section 7.

While computing optimal minorizations poses a significant challenge, a key observation is that one possible choice for φ is $P_{z\cdot} \triangleq (P_{zy} : y \in S)$. With this choice for φ,

$$\frac{K(n)}{n} \to \sum_{x \in S} \pi_x \min_{y \in S} \frac{P_{xy}}{P_{zy}} \quad \text{a.s.} \tag{6}$$

as $n \to \infty$, which is clearly greater than or equal to π_z. In particular, if $X = (X_n : n \ge 0)$ is a sequence of independent and identically distributed random variables (so that all rows of P are identical), then the right-hand side of Eq. (6) is equal to 1, so that regeneration occurs in every transition. While this example is clearly an extreme case, it suggests that when the degree of "row similarity" in P is high, the randomized regenerations associated with the minorization $(\lambda^*, P_{z\cdot})$ can significantly accelerate the rate of regeneration relative to the classical approach.

One possible problem with the use of such randomized regenerations is the potential need to explicitly generate variates from the distributions $(Q_{xy} : y \in S)$ for $x \in S$. This could, for example, require substantial additional programming effort on the part of the simulationist. Fortunately, this can be avoided by using an acceptance–rejection approach due to Glynn and L'Ecuyer [13] and Mykland, Tierney, and Yu [19]. This method permits X to be simulated in an arbitrary fashion (e.g., most discrete time Markov chains are simulated by taking advantage of the fact that they arise as solutions of stochastic recursions of the form $X_{n+1} = g(X_n, Z_{n+1})$). To determine whether a regeneration has occurred at time $n + 1$, one generates a Bernoulli random variable with parameter $\lambda_{X_n} \varphi_{X_{n+1}} / P_{X_n, X_{n+1}}$. If the Bernoulli random variable takes on the value 1, then a regeneration has indeed occurred; otherwise not. This approach to generating X and identifying regenerations is stochastically identical to the method described at the beginning of this section.

3. IMPLICATIONS FOR STEADY-STATE SIMULATION

In this section we will describe the impact of accelerated regeneration on the performance of the regenerative method for steady-state simulation. Building on the framework of Section 2, let f be a real-valued function on S. Our goal is to estimate the steady-state mean $\alpha \triangleq E_\pi[f(X_0)]$, based on simulation of $X_0, X_1, \ldots, X_n$. (If μ is a probability distribution on S, then $E_\mu(\cdot)$ is the expected value operator given that X_0 has distribution μ and P_μ is the corresponding probability measure.)

Suppose that $T(0), T(1), \ldots$, is a sequence of either classical regeneration times (based on returns to state z) or randomized regeneration times based on

a minorization (λ, φ). Set $T(-1) = 0$, and recall that $K(n) = \max\{k \geq -1 : T(k) \leq n\}$ is the index of the cycle ending just prior to time n. Put

$$Y_j = \sum_{k=T(j-1)}^{T(j)-1} f(X_k)$$

and

$$\tau_j = T(j) - T(j-1),$$

for $j \geq 1$. Then the ratio formula for regenerative processes gives

$$\alpha = \frac{E[Y_1]}{E[\tau_1]}.$$

We can now use the regenerative method of steady-state simulation to estimate α (see, e.g., Crane and Lemoine [6], Iglehart and Shedler [15]). When $K(n) \geq 1$, the point estimator for α is

$$\alpha_n \triangleq \frac{\sum_{j=1}^{K(n)} Y_j}{\sum_{j=1}^{K(n)} \tau_j} = \frac{\sum_{j=T(0)}^{T(K(n))-1} f(X_j)}{T(K(n)) - T(0)}.$$

It is well known that α_n obeys a central limit theorem (CLT); namely, there exists σ such that

$$n^{1/2}(\alpha_n - \alpha) \Rightarrow \sigma N(0,1) \tag{7}$$

as $n \to \infty$ (see, e.g., Iglehart and Shedler [15]). The constant σ^2 plays an important role in the construction of confidence intervals for α. In particular, if v_n is a consistent estimator of σ^2 (so that $v_n \Rightarrow \sigma^2$ as $n \to \infty$), then

$$[\alpha_n - z\sqrt{v_n/n}, \alpha_n + z\sqrt{v_n/n}]$$

is an asymptotic $100(1-\delta)\%$ confidence interval for α, provided that z is selected so that $P(N(0,1) \leq z) = 1 - \delta/2$ and $\sigma^2 > 0$. The regenerative method exploits the cycle structure of X to construct such a consistent estimator v_n. In particular, when $K(n) \geq 1$, let

$$v_n = \frac{1}{T(K(n)) - T(0)} \sum_{j=1}^{K(n)} (Y_j - \alpha_n \tau_j)^2.$$

Then, v_n consistently estimates σ^2.

To study the efficiency of the estimator v_n, we need a CLT for v_n. Glynn and Iglehart [12] derived a joint CLT for the point estimator α_n and the standard deviation estimator $s_n = \sqrt{v_n}$. In particular, they derive the following CLT for s_n:

$$n^{1/2}(s_n - \sigma) \Rightarrow N\left(0, \frac{\eta^2}{4\sigma^2}\right), \tag{8}$$

where

$$\eta^2 \triangleq \frac{E[(A_1 - \xi Z_1)^2]}{E[\tau_1]}, \tag{9}$$

and for $i \geq 1$, $Z_i \triangleq Y_i - \alpha\tau_i$, $A_i \triangleq Z_i^2 - \sigma^2\tau_i$, and $\xi \triangleq 2E[Z_1\tau_1]/E[\tau_1]$. Note that

$$n^{1/2}(v_n - \sigma^2) = n^{1/2}(s_n - \sigma)(s_n + \sigma). \tag{10}$$

By Eqs. (8) and (10), we see that v_n satisfies the following CLT:

$$n^{1/2}(v_n - \sigma^2) \Rightarrow N(0, \eta^2), \tag{11}$$

where η^2 is given in Eq. (9). It turns out that the asymptotic variance σ^2 in the CLT for α_n given in Eq. (7) does not depend on whether $T(0), T(1), \ldots,$ are classical or accelerated regeneration times. Similarly, the covariance term in the joint CLT for α_n and s_n does not depend on the choice of the regeneration times $T(0), T(1), \ldots$ (see Andradóttir, Calvin, and Glynn [1]). However, the asymptotic variance η^2 in the CLTs for s_n and v_n given in Eqs. (8), (9), and (11) depends on the choice of regeneration times. To assess the impact of accelerated regeneration on the regenerative method, it is therefore natural to consider the asymptotic variance constant η^2.

Let η_c^2 and η_r^2 be the variance constants that appear on the right-hand side of Eq. (11) corresponding to using classical regeneration (based on returns to z) and randomized regenerations (based on the minorization (λ, φ)), respectively. One might hope that so long as $\sum_x \pi_x \lambda_x \geq \pi_z$ (so that the mean cycle length associated with randomized cycles is less than or equal to that for the classical approach), then $\eta_r^2 \leq \eta_c^2$. Unfortunately, in general, there is no clear-cut relationship between mean cycle length and η^2 (for details, see Glynn and Iglehart [12], Calvin [5]). However, there is one important special case in which something can be said. We will argue that when we choose $\varphi = (P_{zy} : y \in S)$ and $\lambda = \lambda^*$; then we can in fact conclude that $\eta_r^2 \leq \eta_c^2$.

Let $T_c(k)(T_r(k))$ be the time of the kth classical regeneration (randomized regeneration). Set $T_c'(k) = T_c(k) + 1$ for $k \geq 0$, and note that we can effectively view either $T_c'(k)$ or $T_c(k)$ as the time of the kth classical regeneration because, in particular, the cycle quantities $\sum_{j=T_c(k-1)}^{T_c(k)-1} f(X_j) = \sum_{j=T_c'(k-1)}^{T_c'(k)-1} f(X_j)$ and $T_c(k) - T_c(k-1) = T_c'(k) - T_c'(k-1)$ are identical (note that $f(X_{T_c(k-1)}) = f(X_{T_c'(k)-1})$ for all k). The nice feature of the $T_c'(k)$'s is that they form a subsequence of the $T_r(k)$'s, because $\lambda_z^* = 1$. We can now invoke a general result about regenerative steady-state simulations.

Suppose that $\mathfrak{X} = (\mathfrak{X}(t) : t \geq 0)$ is a real-valued stochastic process with two associated increasing sequences of random times $(T_1(k) : k \geq 0)$ and $(T_2(k) :$

$k \geq 0$) such that $(T_2(k): k \geq 0)$ is a subsequence of $(T_1(k): k \geq 0)$. (To handle discrete-time sequences $(\mathfrak{X}_n: n \geq 0)$, set $\mathfrak{X}(t) = \mathfrak{X}_{\lfloor t \rfloor}$.) Let $\theta(t) = I(t \in \{T_2(k): k \geq 0\})$ be a process that "marks" those t that are "type 2" event epochs. Assume that

$\mathfrak{X}$ is a nondelayed regenerative process with respect to $(T_2(k): k \geq 0)$, and $\tilde{\mathfrak{X}}$ is a nondelayed regenerative process with respect to $(T_1(k): k \geq 0)$, where $\tilde{\mathfrak{X}}(t) = (\mathfrak{X}(t), \theta(t))$.

For $i \in \{1,2\}$ and $j \geq 1$, let

$$Y_{ji} = \int_{T_i(j-1)}^{T_i(j)} \mathfrak{X}(s)\, ds,$$

$$\tau_{ji} = T_i(j) - T_i(j-1),$$

and set

$$\alpha = \frac{E[Y_{ji}]}{E[\tau_{ji}]}, \quad Z_{ji} = Y_{ji} - \alpha\tau_{ji}, \quad \text{and} \quad \sigma^2 = \frac{E[Z_{ji}^2]}{E[\tau_{ji}]}.$$

Also, for $i \in \{1,2\}$ and $t \geq 0$, let $K_i(t) = \max\{n \geq 0: T_i(n) \leq t\}$, and when $K_i(t) \geq 1$, let $\alpha_i(t) = \sum_{j=1}^{K_i(t)} Y_{ji} / \sum_{j=1}^{K_i(t)} \tau_{ji}$ and $v_i(t) = [1/T(K_i(t))] \sum_{j=1}^{K_i(t)} (Y_{ji} - \alpha_i(t)\tau_{ji})^2$ be the regenerative point and variance estimators, respectively, based on type-i regenerations, available after t units of computer time have been expended. Put $p = P(T_1(1) = T_2(1))$ and $q = 1 - p$.

THEOREM 1: *If* $E\left[\left(\int_0^{T_2(1)} (|\mathfrak{X}(s)| + 1)\, ds\right)^4\right] < \infty$, *then, for* $i = 1,2$,

$$t^{1/2}(\alpha_i(t) - \alpha) \Rightarrow \sigma N(0,1) \tag{12}$$

and

$$t^{1/2}(v_i(t) - \sigma^2) \Rightarrow \eta_i N(0,1) \tag{13}$$

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

$$\eta_2^2 = \eta_1^2 + 4\sigma^2 \left\{ \frac{q}{p} E\left[Z_{11}^2 \middle| T_2(1) > T_1(1)\right] + \left(\frac{q}{p}\right)^2 \left(E\left[Z_{11} \middle| T_2(1) > T_1(1)\right]\right)^2 \right\}. \tag{14}$$

Theorem 1 states that when one sequence of regeneration times is a "supersequence" of another the regenerative variance estimator based on the supersequence is better (in the sense that its asymptotic variance is smaller). Setting $\mathfrak{X}(t) = f(X_{\lfloor t \rfloor})$, Theorem 1 implies that accelerated regeneration, based on $\varphi = (P_{zy}: y \in S)$ and λ^*, improves the quality of the regenerative variance estimator relative to classical regeneration. Note that the difference between the quality of the regenerative variance estimators obtained through classical and

accelerated regeneration, respectively, increases as $p = P(T_1(1) = T_2(1))$ decreases and as $E\left[Z_{11}^2 \middle| T_2(1) > T_1(1)\right]$ and $\left|E\left[Z_{11} \middle| T_2(1) > T_1(1)\right]\right|$ increase.

In addition, consider two minorizations (λ_1, φ) and (λ_2, φ) with $\lambda_1 \geq \lambda_2$. Suppose that the method of common random numbers is used to generate the Bernoulli sequences for the two associated families of randomized regenerations. Then, it follows that the randomized regenerations corresponding to (λ_2, φ) will be a subsequence of those corresponding to (λ_1, φ). Theorem 1 then implies that the variance estimator corresponding to (λ_1, φ) will have a lower asymptotic variance than the variance estimator corresponding to (λ_2, φ). Consequently, for a given φ, choosing $\lambda = \lambda^*$ is the best possible choice of λ from the standpoint of variance estimation.

Accelerated regeneration also has implications for "low bias" estimation of the steady-state mean α. It is well known that there typically exists a nonzero constant b such that

$$E\alpha_n = \alpha + \frac{b}{n} + o(1/n)$$

as $n \to \infty$. Meketon and Heidelberger [17] established, under certain regularity conditions, that when $T(0) = 0$, then

$$E\alpha_{T(K(n)+1)} = \alpha + o(1/n)$$

as $n \to \infty$, thereby reducing the asymptotic bias. Note, however, that the "low bias" estimator $\alpha_{T(K(n)+1)}$ requires simulating the chain X up to the end of the cycle in progress at time n. Clearly, the additional computer time is reduced when one passes to a new regenerative sequence that is a supersequence of the original one. In particular, efficiencies are realized by using accelerated regeneration.

The efficiency improvement is accentuated in the parallel computing environment, in which multiple parallel processors independently simulate the process to time $T(K(n) + 1)$. (In the parallel context, the effect of bias is magnified, so it is especially important to pass from α_n to $\alpha_{T(K(n)+1)}$.) The completion time, at which the last processor completes its run, occurs at the maximum of the $T(K(n) + 1)$'s collected over all processors. Because of "length-biasing," the tail of the random variable $T(K(n) + 1)$ is fat compared to a typical cycle τ_1, and the impact of this fat tail on the distribution of the maximum can be profound (for details see Glynn and Heidelberger [11]). The point here is that when one uses accelerated regeneration to construct a new sequence of regeneration times that is a supersequence of the original ones, the random variable $T(K(n) + 1)$ is reduced and the tail probabilities that impact completion time are made smaller. Thus, accelerated regeneration can have a particularly positive effect in the parallel computing context.

4. EXTENSIONS TO GENERAL STATE SPACE

As indicated in Section 2, the original motivation for considering randomized regenerations lies in its application to general state space Markov chains. Such chains arise naturally in the study of discrete-event dynamical systems (for details see, e.g., Glynn [9]). In addition, explicit use of randomized regeneration in the simulation context has occurred in two distinct applications: simulation of TES processes (see Asmussen and Melamed [2]) and Gibbs sampling algorithms for computing Bayesian posterior distributions (see Mykland et al. [19]).

In this section we will establish a few basic facts and describe some ideas that are pertinent to randomized regeneration in the general state space setting. We start by letting $X = (X_n : n \geq 0)$ be a Markov chain taking values in a complete separable metric space S. Let $\mathcal{T}$ denote the Borel σ-algebra associated with S. We will assume that X is a Harris chain. This is equivalent to requiring that there exist an integer $m \geq 1$, a probability distribution φ_m, a scalar $\epsilon > 0$, and a nonnegative (measurable) function λ_m such that

(a) $P(X_m \in \cdot \,|\, X_0 = x) \geq \lambda_m(x)\varphi_m(\cdot)$, $\forall x \in S$, and

(b) $P(\lambda_m(X_n) \geq \epsilon \text{ infinitely often } | X_0 = x) = 1$, $\forall x \in S$.

Such chains automatically possess a nontrivial σ-finite invariant measure π that is unique up to a multiplicative constant. We shall require that X be positive recurrent, so that we can assume, without loss of generality, that π is a probability on S.

We start with a definition.

DEFINITION 2: *We say that* (λ_m, φ_m) *is an* m-minorization *of* X *if parts* (a) *and* (b) *hold.*

Note that for each $x \in S$, the measure $\lambda_m(x)\varphi_m(\cdot)$ is absolutely continuous with respect to $P(X_m \in \cdot \,|\, X_0 = x)$, so the Radon–Nikodym theorem guarantees the existence of a function $w_m(x, y)$ such that

$$\lambda_m(x)\varphi_m(dy) = w_m(x, y)P(X_m \in dy \,|\, X_0 = x)$$

for $x, y \in S$ (for details see Royden [22]). Assume for the moment that X possesses a 1-minorization (λ_1, φ_1). Assuming that X is generated arbitrarily and that an associated Bernoulli sequence, having parameters $(w_1(X_n, X_{n+1}) : n \geq 0)$, is simulated, a sequence of randomized regenerations for X can then be constructed as in the discrete state space setting. The regenerations are epochs at which X is distributed according to φ_1, independently of the past. It is easy to show that if $K(n)$ is the number of regenerations occurring in $[0, n]$, then

$$\frac{K(n)}{n} \to \int_S \lambda_1(x)\pi(dx) \quad \text{a.s.}$$

as $n \to \infty$. Therefore, just as in the discrete state space setting, one wishes to choose λ_1 as large as possible. Note that $P(X_1 \in \cdot \,|\, X_0 = x)$ can be written as the sum of a measure that is absolutely continuous with respect to φ_1 and a singular component $R_1(x, \cdot)$, so that there exists a function $r_1(x, \cdot)$ for which

$$P(X_1 \in dy \,|\, X_0 = x) = r_1(x, y)\varphi_1(dy) + R_1(x, dy)$$

for $x, y \in S$. The desire to choose λ_1 as large as possible suggests choosing, for a given φ_1,

$$\lambda_1^*(x) = \inf\{\nu : \varphi_1(\{y : r_1(x, y) < \nu\}) > 0\}$$

for all $x \in S$. (Note that with this definition of $\lambda_1^*(x)$, $\lambda_1^*(x)$ is the largest value of ν such that $\varphi_1(\{y : r_1(x, y) < \nu\}) = 0$ for all $x \in S$.) Theorem 1 can be applied to the general state space setting, assuming that the fourth moment hypothesis demanded there is in force. It guarantees that for a given φ_1 the asymptotic variance of the randomized regenerative variance estimator is minimized by choosing $\lambda_1 = \lambda_1^*$.

A natural question to ask at this juncture is whether there exists a measure φ_1^* that maximizes $\int_S \lambda_1^*(x)\pi(dx)$ over all possible probability measures φ_1. This, of course, is the general state space analog of asking whether finite-dimensional optimization problem (5) always has a solution. The following theorem establishes an affirmative answer to this question.

THEOREM 2: *For all* $m \geq 1$, *there exists an* m-*minorization* $(\lambda_m^*, \varphi_m^*)$ *that maximizes* $\int_S \lambda_m(x)\pi(dx)$ *over all possible* m-*minorizations* (λ_m, φ_m).

Of course, computing $(\lambda_m^*, \varphi_m^*)$ explicitly is, in general, probably not realistic.

We turn now to a discussion of randomized regeneration under an m-minorization with $m > 1$. This is an important case to consider, because there exist Harris chains that arise in applications for which no 1-minorization exists. (A simple theoretical example of such a chain is the process $Y_n = (X_n, X_{n+1})$ constructed from a Harris chain X for which no point is recurrent.) It turns out that in this setting the randomized stopping times that we will construct will no longer divide the sample path of X into independent and identically distributed cycles. However, a weakened form of regeneration will ensue.

There are two slightly different approaches to constructing such "weak regenerations" in the $m > 1$ setting. In both approaches X can be simulated arbitrarily. The first method then generates a sequence of Bernoulli random variables with parameter $(w_m(X_{mk}, X_{m(k+1)}) : k \geq 0)$. Given that the kth Bernoulli is 1, it follows that $X_{m(k+1)}$ is distributed according to φ_m, independent of the position of X_{mk}. This implies that the process evolution up to time mk is independent of the process evolution subsequent to time $m(k+1) - 1$ but that no conclusions can be reached about the dependency situation of the intermediate values $X_{mk+1}, \ldots, X_{m(k+1)-1}$. Consequently, if the $T(i)$'s are the consecutive times at which X is distributed according to φ_m (as indicated by a "successful" Bernoulli coin flip), the cycles $\mathcal{C}_i \triangleq (X_{T(i-1)}, \ldots, X_{T(i)-1})$ are no longer independent, although they are identically distributed. However, it is clear that the cycle $\mathcal{C}_i$ is independent of $\mathcal{C}_j$ for $|i - j| \geq 2$, because such cycles

are automatically separated by at least m time units. This is known as 1-dependent regenerative cycle structure (see, e.g., Glynn [8]). A discussion of steady-state confidence interval methodology for 1-dependent regenerative processes can be found in Glynn [10].

One possible difficulty with the method described above occurs when X is a periodic Harris chain with period $p > 1$. If p and m are not relatively prime, then it is possible that the epochs on which $\lambda_m(X_n) \geq \epsilon$ may not occur along integer multiples of m, due to the periodic structure of X. (It should be noted that the minorization integer m has nothing to do with the period of the chain; they are typically unrelated.) In such a case, it could occur that only finitely many $T(k)$'s can be constructed from X. On the other hand, if X is aperiodic, then when this first method for constructing randomized regeneration times is used, one can show that

$$\frac{K(n)}{n} \to \frac{1}{m} \int_S \lambda_m(x)\pi(dx) \quad \text{a.s.}$$

as $n \to \infty$. Thus, Theorem 2 shows that there exists an optimal m-minorization in this context.

The second method avoids this difficulty and uses the idea of a "regeneration subset." Let $A \subset S$ be a subset such that $\inf\{\lambda_m(x) : x \in A\} > 0$ and for which $P(X_n \in A \text{ infinitely often } | X_0 = x) = 1$ for all $x \in S$; such subsets always exist for Harris chains (see, e.g., Orey [21]). Then we again can generate X arbitrarily. Let

$$T_0(A) = \inf\{n \geq 0 : X_n \in A\}, \qquad T_k(A) = \inf\{n \geq T_{k-1}(A) + m : X_n \in A\}$$

$(k \geq 1)$ be the consecutive visits to A, separated by at least m time units. We now generate a sequence of Bernoulli random variables with corresponding parameters

$$(w_m(X_{T_k(A)}, X_{T_k(A)+m}) : k \geq 0).$$

As in the first method, the epochs $T_k(A) + m$ at which the Bernoulli random variables are equal to 1 form 1-dependent regeneration times for X. A natural question to ask here is how the frequency of "regeneration" of the two variants compare (assuming both methods apply). No easy answer to this question appears possible. The type of regenerative structure constructed with either of these two methods turns out to be a special case of what is known, in the literature, as "wide-sense regeneration" (for the definition and details, see Thorisson [24]).

Note that with either method the cycles constructed are no longer independent. Perhaps surprisingly, it is possible to effectively convert the preceding 1-dependent simulations into a conventional regenerative simulation with *independent and identically distributed* cycles. For either of the two preceding variants, rather than continue simulating the chain from the time of the first

regeneration, one can generate a new $X_0 \sim \varphi_m$, and independently simulate a new path to the time of the next regeneration; then repeat the process. Now, one gets independent cycles, but the joint distributions of this newly simulated process $\tilde{X} = (\tilde{X}_n : n \geq 0)$ are no longer identical to those of the original chain. However, the new process $\tilde{X}$ is *marginally identical* to the chain X in the sense that for each $n \geq 0$, $\tilde{X}_n \overset{\mathcal{D}}{=} X_n$ ($\overset{\mathcal{D}}{=}$ denotes "equality in distribution"). As a consequence, $\tilde{X}$ and X share the same limiting distribution (see Glynn [10]). The process with independent cycles may give rise to point estimators for limiting quantities that are either more variable or less variable than those associated with simulating the original chain. In particular, the variance constants entering the respective CLTs are typically different.

We conclude this section with additional discussion of the m-minorization $(\lambda_m^*, \varphi_m^*)$ that maximizes

$$\int_S \lambda_m(x)\pi(dx)$$

when m is large. Recall that when X is an aperiodic Harris chain,

$$\| P(X_m \in \cdot \,| X_0 = x) - \pi(\cdot) \| \to 0$$

as $m \to \infty$, where $\| \cdot \|$ denotes the total variation norm (see Meyn and Tweedie [18]). This would appear to suggest that for large m, the probability $P(X_m \in \cdot \,| X_0 = x)$ is "similar" to π, so that the function $\lambda_m^*(\cdot)$ associated with $\varphi_m = \pi$ ought to converge to 1. Consequently, choosing $\varphi_m = \pi$ ought to be approximately optimal for large m.

Unfortunately, this heuristic argument is, in general, not valid. Consider, for example, the chain X on state space $Z^+ = \{0,1,2,\dots\}$ for which $P_{x,0} = \frac{1}{2} = P_{x,x+1}$ for $x \geq 0$. This chain is positive recurrent (in fact, uniformly ergodic) with $\pi_x > 0$ for $x \geq 0$. However, observe that if $\varphi_m = \pi$, then $P^m_{x,x+m+1}/\pi_{x+m+1} = 0$ for all $m \geq 1$, so that $\lambda_m^*(x) = 0$ for all $x \in S$. It follows that choosing $\varphi_m = \pi$ is not asymptotically optimal for this chain.

It is worth noting that virtually all the ideas of the section carry over to regenerations constructed from minorizations in which the integer m appearing in condition (a) is permitted to depend on x. However, it is unclear that this generalization broadens the practical scope of our acceleration technique in any significant way.

5. MAXIMAL REGENERATION

An important idea in Markov chain theory is that of "coupling." The idea is to define, on a common probability space, both the original Markov process and a stationary version of that process. One defines the joint probability distribution in such a way that the two chains will eventually occupy the same state at the same time; such a random time is called a "coupling time." Such coupling times provide information on the rate of convergence of the original Markov

chain $X = (X_n : n \geq 0)$ to its stationary distribution. In particular, if κ is the coupling time, and π is the stationary distribution, then

$$\|P(X_n \in \cdot) - \pi(\cdot)\| \leq 2P(\kappa > n). \tag{15}$$

Thus, the tail of the coupling time distribution gives an upper bound on the rate of convergence to steady state. It turns out that there exists a "maximal coupling," for which equality occurs in Eq. (15). For additional details, see Lindvall [16].

One might hope for a similarly tight connection between regeneration times and rates of convergence to steady state. In particular, one might hope for an inequality like Eq. (15), in which κ is replaced by the regeneration time τ_1. Unfortunately, such a result cannot hold in general. For example, consider the (s, S) inventory model in which X evolves according to the stochastic recursion

$$X_{n+1} = \begin{cases} X_n - D_{n+1}, & X_n - D_{n+1} > s, \\ S, & X_n - D_{n+1} \leq s, \end{cases}$$

where the sequence of demands $D = (D_n : n \geq 1)$ is independent and identically distributed. If there exists $\epsilon > 0$ such that $P(D_n \geq \epsilon) = 1$, then the hitting time of state S is clearly bounded by $(S - s)/\epsilon$, uniformly in the initial state. It follows that the tail of the regeneration time distribution is identically zero far enough out. However, the total variation distance of X_n's distribution from π is positive for all $n \geq 1$.

On the other hand, we note that the tail of the hitting time of any state z clearly goes to 0 more slowly than the total variation distance for any Markov chain describing a sequence of independent and identically distributed random variables. Thus, the tail of the regeneration time can go to 0 both faster and slower than the rate of convergence to steady state.

By weakening our mathematical requirements, however, we can establish at least a loose connection between the two concepts. Throughout the remainder of this section, we assume that $X = (X_n : n \geq 0)$ is a finite-state aperiodic irreducible Markov chain with stationary distribution π.

DEFINITION 3: *We say that* γ^{-1} *is the* relaxation time *of* X *if*

$$\gamma = -\limsup_{m \to \infty} \sup_{x \in S} \frac{1}{m} \log(\|P(X_m \in \cdot \,|\, X_0 = x) - \pi(\cdot)\|).$$

We note that the relaxation time is 0 for a Markov chain that describes an independent and identically distributed sequence of random variables and is guaranteed to be nonnegative and finite under our conditions (see Theorem 3). The quantity γ describes the exponential rate at which the total variation distance of X_n's distribution from π converges to 0.

We will also need the following concept.

DEFINITION 4: *Let* τ *be a nonnegative integer valued random variable. The* tail exponent *of* τ *is defined by*

$$\nu = -\limsup_{n \to \infty} \frac{1}{n} \log(P(\tau > n)).$$

Observe that $P(\tau > n)$ is then roughly of the order of magnitude of $e^{-\nu n}$, thereby justifying the terminology "tail exponent."

Let β denote the supremum of all tail exponents of random variables τ_1 that can be expressed as $\tau_1 = T_1 - T_0$, where the T_i's are wide-sense regeneration times for X. We refer to β as the maximal regenerative tail exponent of X. Our main result establishes a relationship between maximal regeneration and the relaxation time. It shows that any irreducible aperiodic finite-state Markov chain possesses a wide-sense regenerative structure with tail behavior reflective of the rate of convergence to steady state.

THEOREM 3: *Let* β *be the maximal regenerative tail exponent of* X, *and let* γ^{-1} *be the relaxation time for* X. *Then,* $\beta \geq \gamma > 0$.

This theorem shows that, at least in principle, the acceleration technique described in this paper can be used to construct regeneration times with tail behavior at least as good as that of the rate of convergence to stationarity.

6. EXTENSIONS TO CONTINUOUS TIME

Suppose that $Y = (Y_n : n \geq 0)$ is a discrete-time Markov chain taking values in a state space S. Assume that for each $x \in S$, there is specified a distribution function $F(x, \cdot)$ corresponding to the lifetime of a nonnegative random variable. We may now construct a continuous-time stochastic process $X = (X(t) : t \geq 0)$ as follows:

$$X(t) = \sum_{j=0}^{\infty} Y_j I\left(\sum_{k=0}^{j-1} \zeta_k \leq t < \sum_{k=0}^{j} \zeta_k\right)$$

where

$$P(\zeta_0 \in dt_0, \ldots, \zeta_k \in dt_k \,|\, Y) = \prod_{j=0}^{k} F(Y_j, dt_j).$$

Such a process X is a semi-Markov process; continuous-time Markov chains form a subclass of this family of processes. Our discrete-time construction of accelerated regenerations suggests that we consider minorizations associated with the transition function of X. However, even when X is a continuous-time Markov chain, this approach is generally impractical, because computing the transition function $P(X(t) = \cdot \,|\, X(0) = x)$ typically requires solving a system of differential equations. (In the general semi-Markov setting, a system of integro-differential equations must be solved.)

Instead, we recommend that one construct minorizations for the discrete-time Markov chain Y. If a 1-minorization (λ, φ) for Y can be constructed, then the epochs at which X makes a state transition distributed according to φ will be regenerations for X yielding independent and identically distributed cycles. If an m-minorization (λ_m, φ_m) for Y is used instead (with $m \geq 2$), the corresponding epochs at which X is distributed according to φ_m continue to yield 1-dependent cycle structure. (However, it turns out that the regenerations now constructed are no longer "wide-sense regenerative" for X, in the sense of Thorisson [24].)

While the principle of discrete-time conversion (see Fox and Glynn [7]) can be used to avoid the need for continuous-time regeneration for estimation of steady-state quantities, there is no way to avoid such a need in dealing with estimation of transient quantities; such applications are described in Section 8.

7. EXAMPLES

We return to nonlinear program (5):

$$\max_{\varphi} \sum_{x \in S} \pi_x \min_{y \in S} \frac{P_{xy}}{\varphi_y}$$

$$\text{subject to:} \quad \sum_{y \in S} \varphi_y = 1, \quad \text{and} \quad \varphi_y \geq 0, \quad \text{for all } y \in S,$$

and suppose that $S = \{1,2,\ldots,N\}$. Assume first that $P_{xy} > 0$ for all $x, y \in S$. Let $h: S \to S$, and let $A(h)$ denote the subset of probabilities φ on S such that

$$\frac{P_{xh(x)}}{\varphi_{h(x)}} \leq \frac{P_{xy}}{\varphi_y} \quad \forall x, y.$$

Then $A(h)$ is a convex set for each h, and $\bigcup_h A(h)$ contains all probabilities on S. On $A(h)$

$$F(\varphi) \triangleq \sum_{x \in S} \pi_x \min_{y \in S} \frac{P_{xy}}{\varphi_y} = \sum_{x \in S} \pi_x \frac{P_{xh(x)}}{\varphi_{h(x)}}$$

is strictly convex. (Note that if $\varphi \in A(h) \cap A(h')$, then $F(\varphi)$ is the same for either representation.)

We claim that the optimal φ must lie in the intersection of a maximal collection of $A(h)$'s and that this intersection is a singleton. If $\varphi \in \bigcap_{i \in I} A(h_i)$ and $\bigcap_{i \in I} A(h_i)$ is not a singleton, then, because F is strictly convex, φ must be an extreme point of the intersection, i.e., an extreme point of the set

$$\left\{ \varphi : \frac{P_{xh_i(x)}}{\varphi_{h_i(x)}} \leq \frac{P_{xy}}{\varphi_y} \quad \forall x, y \in S, \quad \forall i \in I \right\},$$

viewed as a subset of the N-dimensional simplex. This set has two types of extreme points; either at least one of the φ_y's becomes 0, or for at least one x and one $y \notin \{h_i(x) : i \in I\}$, P_{xy}/φ_y becomes minimal for row x. The first type of extreme point cannot be optimal. To see this, suppose that $\varphi_1 = 0$. Let $g(\epsilon) = F((\epsilon, (1-\epsilon)\varphi_2, \ldots, (1-\epsilon)\varphi_N))$. For small enough positive ϵ, $g(\epsilon)$ is increasing. This implies that $\varphi_1 = 0$ cannot hold for the optimal φ. Therefore, the optimal φ is one of the second type of extreme point, so it must be that there is some $l: S \to S$, $l \neq h_i$ for $i \in I$, such that for all $i \in I$

$$\frac{P_{xh_i(x)}}{\varphi_{h_i(x)}} \leq \frac{P_{xy}}{\varphi_y} \quad \text{and} \quad \frac{P_{xl(x)}}{\varphi_{l(x)}} \leq \frac{P_{xy}}{\varphi_y} \quad \forall x, y.$$

But then $\varphi \in A(l) \cap \left(\bigcap_{j \in I} A(h_i)\right)$, which establishes the claim.

For $N = 2$, this implies that the optimal φ is one of the rows of P. To see this, represent probabilities on $S = \{1,2\}$ by the mass φ_1 on 1. Without loss of generality suppose that $P_{11} \leq P_{21}$. For $\varphi_1 \in [0, P_{11}]$, F is strictly convex. (Here $[0, P_{11}] = A(h)$, where $h(1) = h(2) = 2$.) Similarly, F is strictly convex for $\varphi_1 \in [P_{11}, P_{21}]$ and for $\varphi_1 \in [P_{21}, 1]$. The optimal φ must then be a row or degenerate. However, the optimal φ cannot be degenerate. For example, if $\varphi_1 = 1$, then $F(\varphi) = \pi_1$, while if $\varphi_1 = P_{11}$, then $F(\varphi) = \pi_1 + \pi_2 P_{22}/P_{12}$.

In higher dimensions it is not so easy to find the optimal φ. Because of the preceding claim, we know that the optimal φ has a maximal "binding set"

$$B(\varphi) = \left\{ (x, y) \in S \times S : \frac{P_{xy}}{\varphi_y} < \frac{P_{xz}}{\varphi_z} \text{ for all } z \text{ such that } (x, z) \notin B(\varphi) \right\}.$$

Specification of a maximal $B(\varphi)$ uniquely identifies φ as the solution of a system of linear equations and inequalities. In addition, the optimal φ must be such that each column has a binding element; i.e., for each y, $(x, y) \in B(\varphi)$ for some $x \in S$. If this were not the case, say if P_{xy}/φ_y were not minimal in row x for any x, then φ_y could be increased slightly without changing the objective function value, and at the same time all the other φ_z's could be decreased slightly, thus increasing the objective function.

If all elements of P are not strictly positive, then the preceding argument needs slight modification. The definition of $B(\varphi)$ must be changed to include all (x, y) such that $\varphi_y = 0$. The same argument as before results in the conclusion that for φ to be optimal, it must be uniquely determined by a maximal binding set, where now φ could be determined by the requirement that $\varphi_y = 0$ for some y's, in addition to other linear equalities and inequalities.

The problem of determining the optimal φ is a combinatorial optimization problem. By the preceding characterization, we can identify the finite set of potential optima (maximal binding sets). All of these potential optima must then be compared to find the best solution; this method is illustrated in Example 2, later. Note that any row of P is a candidate, because a binding row uniquely determines φ. For $N > 2$, the optimal φ need not be a row of P, however, as will be seen in Example 2.

Example 1: Consider a birth-death chain on the integers; i.e., $P_{xy} = 0$ if $|x - y| \neq 1$. In this case the only possible regenerations are the classical ones. To see this, suppose that λ_{x_1} and λ_{x_2} are both positive. Then the requirement that $P_{xy} \geq \lambda_x \varphi_y$ for all x, y, implies that $\varphi_y > 0$ only if $P_{x_1,y} > 0$ and $P_{x_2,y} > 0$. But the intersection of the supports of $P_{x_1,\cdot}$ and $P_{x_2,\cdot}$ contain at most one state, which implies that φ is a point mass at some state y. This is then equivalent to classical regenerations at state y.

Example 2: Let $(X_n : n \geq 0)$ be a random walk on the integers modulo $N \geq 2$, with

$$P_{xy} = \begin{cases} \frac{1}{2}, & x = y, \\ \frac{1}{2(N-1)}, & x \neq y. \end{cases}$$

The stationary distribution is uniform, and we claim that this is the optimal φ in terms of maximizing the regeneration rate. This claim clearly holds for $N = 2$, as in that case we know the optimal φ is a row of the matrix P, and when $N = 2$ both rows of P correspond to the uniform distribution. Therefore, suppose that $N > 2$. We will explicitly identify all maximal binding sets. Let $B_0 = \{(x, y) \in S \times S : x \neq y\}$; this is the maximal binding set corresponding to the uniform distribution $\varphi_i = 1/N$ for $i = 1, \ldots, N$. It is clear that B_0 is the unique maximal binding set that contains no diagonal element. Next, suppose that $B_x(\varphi)$ is a maximal binding set that contains (x,x). Then we must have that $\varphi_x \geq \frac{1}{2}$ and that column x is binding. We claim that B_x must contain exactly column x and row x. If $\varphi_x > \frac{1}{2}$, then there will be at least one element in row x that is not binding, so B_x is not maximal. We therefore have that $\varphi_x = \frac{1}{2}$ and that $\varphi_z = 1/(2(N-1))$ for $z \neq x$, so φ is the xth row $P_{x\cdot}$ of P. This establishes the claim that the only possibilities for optimal φ are the uniform distribution or a row of P. Note that the λ^*'s corresponding to $\varphi = \pi$ are $\lambda_x^* = N/(2(N-1))$ for all x, and so the regeneration rate converges to $\frac{1}{2}$ as the size of the state space N increases. If φ were chosen to be a row of P, then the regeneration rate would converge to 0 as $N \to \infty$.

Example 3: We next consider a general state space Markov chain, namely the waiting time sequence for the M/M/1 queue with service rate 1 and arrival rate $\rho < 1$. The state space is $[0,\infty)$ and the transition function is given, for $x \geq 0$, by

$$P(x,\{0\}) = \frac{1}{1+\rho} e^{-\rho x},$$

$$P(x,dy) = p(x,y)\,dy = \frac{\rho}{1+\rho} e^{-\rho(x-y)^+ - (y-x)^+}\,dy, \qquad y \geq 0.$$

The stationary distribution π is given by

$$\pi(\{0\}) = 1 - \rho, \qquad \pi(dx) = \rho(1-\rho)e^{-(1-\rho)x}\,dx, \qquad x > 0.$$

If we choose $\varphi(\cdot) = P(0,\cdot)$, we obtain $\lambda^*(0) = 1$, and for $x > 0$,

$$\lambda^*(x) = \min\left(\frac{P(x,\{0\})}{\varphi(\{0\})}, \inf_{y>0} \frac{p(x,y)}{p(0,y)}\right)$$

$$= \min\left(e^{-\rho x}, \inf_{y>0} e^{y-\rho(x-y)^+-(y-x)^+}\right) = e^{-\rho x},$$

so that $\lambda^*(x) = e^{-\rho x}$, for all $x \geq 0$. Then,

$$\int_0^\infty \pi(dx)\lambda^*(x) = 1 - \rho + \int_0^\infty \rho(1-\rho)e^{-(1-\rho)x}e^{-\rho x}\,dx = 1 - \rho^2.$$

We conclude that the expected cycle length with this choice of φ is $(1-\rho^2)^{-1}$. Using classical regeneration with 0 as return state gives a mean cycle length of $(1-\rho)^{-1}$. As $\rho \nearrow 1$, we see that the expected cycle length is about twice as large for the classical regenerative method.

A similar calculation shows that the best φ of the form $\varphi(\cdot) = P(x,\cdot)$ is the one chosen above, i.e., with $x = 0$.

8. APPLICATIONS TO TERMINATING SIMULATIONS

The accelerated regenerations introduced earlier in this paper can also be used to advantage in dealing with certain performance measures that arise in the terminating simulation context. We illustrate this with a couple of examples.

Suppose that we wish to estimate the transient quantity $\alpha = E[f(X_n)]$, where f is a real-valued function defined on the state space of the discrete-time Markov chain $X = (X_n : n \geq 0)$. Assume that πf is known. (This might occur, for example, in dealing with product-form queueing networks.) Then, the quantity α can be computed by taking advantage of the coupling concepts described earlier in this paper. In particular, if $X^* = (X_n^* : n \geq 0)$ is the associated stationary version to which X is coupled, and κ is the corresponding coupling time, then

$$\alpha = \pi f + E[(f(X_n) - f(X_n^*))I(\kappa > n)].$$

Hence, α can be estimated by replicating independent and identically distributed copies of the random variable

$$(f(X_n) - f(X_n^*))I(\kappa > n);$$

this idea is explained in greater detail in Glynn and Wong [14].

One way of coupling Markov chains is to define κ as the first time that the chains X and X^* simultaneously regenerate; this approach is especially conve-

nient in general state space, where it is common to simulate chains that visit no point infinitely often, thereby rendering coupling based on simultaneous visits to a common state impractical. Clearly, the random variable κ decreases when one passes from a given pair of regenerative sequences to ones that form supersequences of the original regenerations. In this case the variance of $(f(X_n) - f(X_n^*))I(\kappa > n)$ is reduced and, furthermore, the computational effort necessary to simulate the pair of processes X and X^* to time $\min\{\kappa, n\}$ is diminished. Thus, accelerated regeneration has obvious value when applied in conjunction with such coupling ideas.

A second illustration of the utility of accelerated regeneration in the terminating simulation context arises when considering the expected "cumulative cost"

$$\alpha = E\left[\sum_{i=0}^{n} f(X_i)\right],$$

where f and X are as above. Again, suppose that πf is known. Then,

$$\alpha - (n+1)\pi f = -E\left[\sum_{i=n+1}^{T(K(n)+1)-1} (f(X_i) - \pi f)\right],$$

where $T(K(n)+1)$ is the epoch of the first regeneration subsequent to time n. (This identity follows from Wald's equality.) Shahabuddin [23] has suggested estimating α on the basis of the preceding identity, using replicates of the random variable

$$\sum_{i=n+1}^{T(K(n)+1)-1} (f(X_i) - \pi f).$$

It seems reasonable that if one can construct accelerated regenerations that form a supersequence of the original regeneration times, then it is preferable to pass to the newly constructed supersequence.

References

1. Andradóttir, S., Calvin, J.M., & Glynn, P.W. (1994). Increasing the frequency of regeneration for Markov processes. In J.D. Tew, M.S. Manivannan, D.A. Sadowski, & A.F. Seila (eds.), *Proceedings of the 1994 Winter Simulation Conference.* New York: Association for Computing Machinery, pp. 320–323.
2. Asmussen, S. & Melamed, B. (1994). Regenerative simulation of TES processes. *Acta Applicandae Mathematicae* 34: 237–260.
3. 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.
4. Billingsley, P. (1968). *Convergence of probability measures.* New York: Wiley.
5. Calvin, J.M. (1994). Return-state independent quantities in regenerative simulation. *Operations Research* 42: 531–542.
6. Crane, M.A. & Lemoine, A.J. (1977). *An introduction to the regenerative method for simulation analysis.* Berlin: Springer-Verlag.

7. Fox, B. & Glynn, P.W. (1986). Discrete-time conversion for simulating semi-Markov processes. *Operations Research Letters* 5: 191–196.
8. Glynn, P.W. (1982). Simulation output analysis for general state space Markov chains. Ph.D. dissertation, Department of Operations Research, Stanford University, Stanford, CA.
9. Glynn, P.W. (1989). A GSMP formalism for discrete event systems. *Proceedings of the IEEE* 77: 14–23.
10. Glynn, P.W. (1994). Some topics in regenerative steady state simulation. *Acta Applicandae Mathematicae* 34: 225–236.
11. Glynn, P.W. & Heidelberger, P. (1991). Analysis of parallel replicated simulations under a completion time constraint. *ACM Transactions on Modeling and Computer Simulation* 1: 3–23.
12. Glynn, P.W. & Iglehart, D.L. (1987). A joint central limit theorem for the sample mean and regenerative variance estimator. *Annals of Operations Research* 8: 41–55.
13. Glynn, P.W. & L'Ecuyer, P. (1995). Likelihood ratio gradient estimation for stochastic recursions. *Advances in Applied Probability* (to appear).
14. Glynn, P.W. & Wong, E. (1996). Efficient simulation via coupling. Technical Report (in preparation).
15. Iglehart, D.L. & Shedler, G.S. 1980. *Regenerative simulation of response times in networks of queues.* Berlin: Springer-Verlag.
16. Lindvall, T. (1992). *Lectures on the coupling method.* New York: Wiley.
17. Meketon, M.S. & Heidelberger, P. (1982). A renewal theoretic approach to bias reduction in regenerative simulations. *Management Science* 28: 173–181.
18. Meyn, S.P. & Tweedie, R.L. (1993). *Markov chains and stochastic stability.* New York: Springer-Verlag.
19. Mykland, P., Tierney, L., & Yu, B. (1992). Regeneration in Markov chain samplers. Technical Report 585, School of Statistics, University of Minnesota.
20. Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 43: 309–318.
21. Orey, S. (1971). *Limit theorems for Markov chain transition probabilities.* London: Van Nostrand Reinhold.
22. Royden, H. (1988). *Real analysis.* New York: Macmillan.
23. Shahabuddin, P. (1994). Efficient simulation of regenerative systems whose steady state measure is known. Research Report 19658, IBM Watson Research Center, Yorktown Heights, NY.
24. Thorisson, H. (1992). Construction of a stationary regenerative process. *Stochastic Processes and Their Applications* 42: 237–253.

APPENDIX

We begin by establishing some results that will be used in the proof of Theorem 1.

LEMMA 1: *Let $Y_{11}(1)$ and $Y_{11}(2)$ ($\tau_{11}(1)$ and $\tau_{11}(2)$) be random variables with the distribution of Y_{11} (τ_{11}) conditional on $\{T_1(1) = T_2(1)\}$ and $\{T_1(1) < T_2(1)\}$, respectively. Set $Z_{11}(k) = Y_{11}(k) - \alpha\tau_{11}(k)$ for $k = 1,2$. Then, under the assumptions of Theorem 1,*

$$E[Y_{12}] = \frac{1}{p} E[Y_{11}], \tag{16}$$

$$E[Y_{12}^2] = \frac{1}{p} E[Y_{11}^2] + 2\frac{q}{p^2} E[Y_{11}(2)]E[Y_{11}], \tag{17}$$

$$E[Y_{12}^3] = \frac{1}{p} E[Y_{11}^3] + 3\frac{q}{p^2} E[Y_{11}(2)^2]E[Y_{11}] + 3\frac{q}{p^2} E[Y_{11}(2)]E[Y_{11}^2] + 6\frac{q^2}{p^3} (E[Y_{11}(2)])^2 E[Y_{11}], \tag{18}$$

$$E[Z_{12}^4] = \frac{1}{p} E[Z_{11}^4] + 6\frac{q}{p^2} E[Z_{11}(2)^2]E[Z_{11}^2] + 4\frac{q}{p^2} E[Z_{11}(2)]E[Z_{11}^3] + 12\frac{q^2}{p^3} (E[Z_{11}(2)])^2 E[Z_{11}^2]. \tag{19}$$

Notice that Eqs. (16)-(18) remain true with τ or Z replacing Y.

PROOF: For $k > 0$, set

$$\nu_k = \sum_{l=0}^{\infty} I(T_2(l) = T_1(k)) = \theta(T_1(k)).$$

Let

$$\psi(t) = E[e^{itY_{12}}], \qquad \psi_j(t) = E[e^{itY_{11}(j)}], \quad \text{for } j = 1,2, \qquad \text{and} \qquad \xi(t) = E[e^{itY_{11}}].$$

Then,

$$\begin{aligned} E[e^{itY_{12}}] &= \sum_{k=1}^{\infty} E[e^{it(Y_{11} + \cdots + Y_{k1})} I(\nu_1 = \cdots = \nu_{k-1} = 0;\ \nu_k = 1)] \\ &= \sum_{k=1}^{\infty} \prod_{j=1}^{k-1} E[e^{itY_{j1}} I(\nu_j = 0)] E[e^{itY_{k1}} I(\nu_k = 1)] \\ &= \sum_{k=0}^{\infty} (q\psi_2(t))^k (p\psi_1(t)) = p\psi_1(t)(1 - q\psi_2(t))^{-1}. \end{aligned}$$

Therefore,

$$(1 - q\psi_2(t))\psi(t) = p\psi_1(t) = \xi(t) - q\psi_2(t). \tag{20}$$

By the moment hypothesis of the theorem, ψ, ψ_2, and ξ are four times differentiable at 0. By differentiating Eq. (20) n times ($n \le 4$), we obtain the identity

$$E[Y_{12}^n] - q\sum_{k=0}^{n} \binom{n}{k} E[Y_{11}(2)^{n-k}]E[Y_{12}^k] = E[Y_{11}^n] - qE[Y_{11}(2)^n]$$

or

$$pE[Y_{12}^n] = E[Y_{11}^n] + q\sum_{k=1}^{n-1} \binom{n}{k} E[Y_{11}(2)^{n-k}]E[Y_{12}^k].$$

These identities can be solved recursively to obtain the lemma. Note that the same technique can be used with Z in place of Y. ■

COROLLARY 1: *let f_1, f_2 be two real-valued functions defined on S with corresponding Y random variables $Y_{ji}(f_k)$, $k = 1,2$, and the notation otherwise as in Lemma 1. Then, under the assumptions of Theorem 1,*

$$\begin{aligned} E[Y_{12}(f_1)Y_{12}(f_2)] &= \frac{1}{p} E[Y_{11}(f_1)Y_{11}(f_2)] \\ &\quad + \frac{q}{p^2} (E[Y_{11}(f_1)(2)]E[Y_{11}(f_2)] + E[Y_{11}(f_2)(2)]E[Y_{11}(f_1)]), \\ E[Y_{12}(f_1)^2 Y_{12}(f_2)] &= \frac{1}{p} E[Y_{11}(f_1)^2 Y_{11}(f_2)] \\ &\quad + \frac{q}{p^2} (E[Y_{11}(f_1)(2)^2]E[Y_{11}(f_2)] + E[Y_{11}(f_2)(2)]E[Y_{11}(f_1)^2]) \\ &\quad + 2\frac{q}{p^2} (E[Y_{11}(f_1)Y_{11}(f_2)]E[Y_{11}(f_1)(2)] \\ &\qquad + E[Y_{11}(f_1)]E[Y_{11}(f_1)(2)Y_{11}(f_2)(2)]) \\ &\quad + 2\frac{q^2}{p^3} ((E[Y_{11}(f_1)(2)])^2 E[Y_{11}(f_2)] \\ &\qquad + 2E[Y_{11}(f_1)(2)]E[Y_{11}(f_2)(2)]E[Y_{11}(f_1)]). \end{aligned}$$

Notice that the preceding remain true with Z replacing Y.

PROOF: Apply Eqs. (17) and (18) with $f = r_1 f_1 + r_2 f_2$ and equate coefficients in $r_1 r_2$ in the first case and $r_1^2 r_2$ in the second to obtain the result. ■

PROOF OF THEOREM 1: Equations (12) and (13) follow directly from Glynn and Iglehart [12] (see also Eqs. (8) and (10)). It remains to show that Eq. (14) holds. Applying Eqs. (16) and (17) of Lemma 1 with $f = 1$ gives

$$E[\tau_{12}] = \frac{1}{p} E[\tau_{11}]$$

and

$$E[\tau_{12}^2] = \frac{1}{p} E[\tau_{11}^2] + 2\frac{q}{p^2} E[\tau_{11}(2)]E[\tau_{11}].$$

From Eq. (18) of Lemma 1, we have

$$E[Z_{12}^3] = \frac{1}{p} E[Z_{11}^3] + 3\frac{q}{p^2} E[Z_{11}(2)]E[Z_{11}^2],$$

because $E[Z_{11}] = 0$. A similar application of Corollary 1 yields

$$E[\tau_{12}Z_{12}] = \frac{1}{p} E[\tau_{11}Z_{11}] + \frac{q}{p^2} E[Z_{11}(2)]E[\tau_{11}]$$

and

$$E[\tau_{12}Z_{12}^2] = \frac{1}{p} E[\tau_{11}Z_{11}^2] + \frac{q}{p^2}(E[Z_{11}(2)^2]E[\tau_{11}] + E[\tau_{11}(2)]E[Z_{11}^2])$$
$$+ 2\frac{q}{p^2} E[\tau_{11}Z_{11}]E[Z_{11}(2)] + 2\frac{q^2}{p^3}(E[Z_{11}(2)])^2 E[\tau_{11}].$$

Using these formulas, together with Eq. (19) and

$$E[Z_{1j}^2] = \sigma^2 E[\tau_{1j}],$$

for $j = 1,2$, in Eq. (9) gives

$$\eta_2^2 E[\tau_{12}] = E[Z_{12}^4] - 2\sigma^2 E[\tau_{12}Z_{12}^2] + \sigma^4 E[\tau_{12}^2] - 4\frac{E[\tau_{12}Z_{12}]}{E[\tau_{12}]} E[Z_{12}^3]$$
$$+ 8\sigma^2 \frac{(E[\tau_{12}Z_{12}])^2}{E[\tau_{12}]}$$
$$= \frac{1}{p} E[Z_{11}^4] + 6\frac{q}{p^2} E[Z_{11}(2)^2]E[Z_{11}^2] + 4\frac{q}{p^2} E[Z_{11}(2)]E[Z_{11}^3]$$
$$+ 12\frac{q^2}{p^3}(E[Z_{11}(2)])^2 E[Z_{11}^2]$$
$$- 2\sigma^2\left(\frac{1}{p} E[\tau_{11}Z_{11}^2] + \frac{q}{p^2}(E[Z_{11}(2)^2]E[\tau_{11}] + E[\tau_{11}(2)]E[Z_{11}^2])\right)$$
$$- 2\sigma^2\left(2\frac{q}{p^2} E[\tau_{11}Z_{11}]E[Z_{11}(2)] + 2\frac{q^2}{p^3}(E[Z_{11}(2)])^2 E[\tau_{11}]\right)$$
$$+ \sigma^4\left(\frac{1}{p} E[\tau_{11}^2] + 2\frac{q}{p^2} E[\tau_{11}(2)]E[\tau_{11}]\right)$$
$$- 4\frac{E[\tau_{11}Z_{11}] + \frac{q}{p} E[Z_{11}(2)]E[\tau_{11}]}{E[\tau_{11}]}\left(\frac{1}{p} E[Z_{11}^3] + 3\frac{q}{p^2} E[Z_{11}(2)]E[Z_{11}^2]\right)$$
$$+ 8\sigma^2 p \frac{\left(\frac{1}{p} E[\tau_{11}Z_{11}] + \frac{q}{p^2} E[Z_{11}(2)]E[\tau_{11}]\right)^2}{E[\tau_{11}]}$$

$$= \frac{1}{p}\left(E[Z_{11}^4] - 2\sigma^2 E[\tau_{11}Z_{11}^2] + \sigma^4 E[\tau_{11}^2] - 4\frac{E[\tau_{11}Z_{11}]}{E[\tau_{11}]} E[Z_{11}^3]\right.$$
$$\left. + 8\sigma^2 \frac{(E[\tau_{11}Z_{11}])^2}{E[\tau_{11}]}\right)$$
$$+ 4\frac{\sigma^2}{p}\left(\frac{q}{p} E[Z_{11}(2)^2]E[\tau_{11}] + \frac{q^2}{p^2}(E[Z_{11}(2)])^2 E[\tau_{11}]\right)$$
$$= \frac{1}{p} E[\tau_{11}]\eta_1^2 + 4\frac{\sigma^2}{p}\left(\frac{q}{p} E[Z_{11}(2)^2]E[\tau_{11}] + \frac{q^2}{p^2}(E[Z_{11}(2)])^2 E[\tau_{11}]\right).$$

Because $E[\tau_{12}] = E[\tau_{11}]/p$, dividing through by $E[\tau_{12}]$ gives the theorem. ■

PROOF OF THEOREM 2: It suffices to prove the result for $m = 1$, and we write λ, φ instead of λ_1, φ_1. For all $x \in S$ and $B \in \mathcal{T}$, let $P(x,B) \triangleq P(X_{n+1} \in B \mid X_n = x)$ be the transition probability for the chain. Because $\lambda(x) \le 1$ for all $x \in S$, it must be that

$$s = \sup\left\{\int \lambda(x)\pi(dx) : (\lambda,\varphi) \text{ is a minorization for some probability } \varphi\right\} \le 1.$$

We may assume that there exists a sequence $(\lambda^{(n)}, \varphi^{(n)})$ of minorizations such that

$$\rho^{(n)} \triangleq \int \lambda^{(n)}(x)\pi(dx) \nearrow s$$

as $n \to \infty$. Now, let $B \in \mathcal{T}$ and observe that

$$\pi(B) = \int P(x,B)\pi(dx) \ge \int \lambda^{(n)}(x)\varphi^{(n)}(B)\pi(dx) = \rho^{(n)}\varphi^{(n)}(B).$$

Because $\rho^{(n)} \to s$ as $n \to \infty$, there exists an integer n_0 such that for $n \ge n_0$, $\rho^{(n)} \ge s/2$, and thus, for $n \ge n_0$,

$$\varphi^{(n)}(B) \le \pi(B)/\rho^{(n)} \le 2\pi(B)/s. \tag{21}$$

We now recall that π is a probability on the Borel sets of a complete separable metric space, and hence π is a tight measure (see, e.g., Billingsley [4, Theorem 1.4]). Thus, for each $\epsilon > 0$, there exists a compact set K_ϵ such that $\pi(K_\epsilon) > 1 - \epsilon$. Letting K_ϵ^c denote the set complement of K_ϵ, we see by Eq. (21) that

$$\varphi^{(n)}(K_\epsilon^c) \le 2\pi(K_\epsilon^c)/s$$

for $n \ge n_0$, and hence $\varphi^{(n)}(K_\epsilon) > 1 - 2\epsilon/s$ for $n \ge n_0$, proving that $\{\varphi^{(n)} : n \ge 1\}$ is a tight family of probability measures. Then, by Prohorov's theorem, the family $\{\varphi^{(n)} : n \ge 1\}$ is relatively compact. Hence, there exists a subsequence $\{n_k\}$ such that

$$\varphi^{(n_k)} \Rightarrow \varphi \tag{22}$$

as $k \to \infty$, where φ is a probability on $\mathcal{T}$. But $(\lambda^{(n_k)}, \varphi^{(n_k)})$ is a sequence of minorizations, so that if $x \in S$ and $B \in \mathcal{T}$, then

$$P(x,B) \ge \lambda^{(n_k)}(x)\varphi^{(n_k)}(B)$$

for all k, and this implies that

$$P(x,B) \geq \limsup_k \lambda^{(n_k)}(x)\varphi^{(n_k)}(B) \geq \limsup_k \lambda^{(n_k)}(x)\liminf_k \varphi^{(n_k)}(B). \quad (23)$$

Let $\lambda(x)$ be the $\mathcal{F}$-measurable function $\limsup_k \lambda^{(n_k)}(x)$, and observe that Eq. (23) implies that

$$P(x,B) \geq \lambda(x)\liminf_k \varphi^{(n_k)}(B) \quad (24)$$

for all $x \in S$ and $B \in \mathcal{F}$. In particular, Eq. (24) holds for open sets $G \in \mathcal{F}$, and thus

$$P(x,G) \geq \lambda(x)\liminf_k \varphi^{(n_k)}(G) \geq \lambda(x)\varphi(G),$$

the last inequality by weak convergence relation (22) (see, e.g., Billingsley [4, p. 12]). Because the Borel σ-algebra $\mathcal{F}$ is generated by the open sets,

$$P(x,B) \geq \lambda(x)\varphi(B)$$

for all $x \in S$ and $B \in \mathcal{F}$. Finally, Fatou's lemma yields

$$s = \limsup_k \int \lambda^{(n_k)}(x)\pi(dx) \leq \int \limsup_k \lambda^{(n_k)}(x)\pi(dx) = \int \lambda(x)\pi(dx),$$

which concludes the proof of Theorem 2. ■

PROOF OF THEOREM 3: The positivity of γ follows from the fact that if P is the transition matrix of X, then

$$P^m - \Pi = (P - \Pi)^m$$

for $m \geq 1$, where Π is the matrix having all rows identical to π. Because $P - \Pi$ is a finite matrix and $P^m - \Pi$ converges to 0, it is evident that all the eigenvalues of $P - \Pi$ have modulus strictly less than 1, thereby guaranteeing that $\gamma > 0$.

It remains to show that $\beta \geq \gamma$. Observe that for each $\epsilon > 0$, there exists m_0 such that for $m \geq m_0$,

$$\sup_{x \in S} \frac{1}{m} \log(\| P(X_m \in \cdot \mid X_0 = x) - \pi(\cdot) \|) \leq -\gamma + 2\epsilon,$$

and hence

$$P_{xy}^m \geq \pi_y - e^{-(\gamma - 2\epsilon)m}$$

for all $x, y \in S$. Now, choose $m_1 \geq m_0$ sufficiently large so that

$$e^{-\epsilon m_1} \leq \min\{\pi_y : y \in S\}.$$

Then,

$$P_{xy}^{m_1} \geq \pi_y(1 - e^{-(\gamma - \epsilon)m_1}),$$

so that

$$\frac{P_{xy}^{m_1}}{\pi_y} \geq 1 - e^{-(\gamma - \epsilon)m_1}.$$

Now, consider the wide-sense regenerations associated with the m_1-minorization (π, λ^*), based on the first of the two wide-sense constructions described in Section 4. Then,

$$\lambda^*(x) \geq 1 - e^{-(\gamma - \epsilon)m_1}.$$

It follows that if $\tau_1 = T_1 - T_0$ is the cycle time associated with this construction, then

$$P(\tau_1 > n) = E\left[\prod_{j=0}^{\lfloor n/m_1 \rfloor - 1} (1 - \lambda^*(X_{jm_1}))\right] \leq \exp(-(\gamma - \epsilon)m_1\lfloor n/m_1 \rfloor),$$

from which we may conclude that τ_1 has a tail exponent no smaller than $\gamma - \epsilon$. But ϵ was arbitrary, so the inequality $\beta \geq \gamma$ ensues. This concludes the proof. ■