Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

A COMPARISON OF OUTPUT-ANALYSIS METHODS FOR SIMULATIONS OF PROCESSES WITH MULTIPLE REGENERATION SEQUENCES

James M. Calvin
Marvin K. Nakayama

Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, U.S.A.

ABSTRACT

We compare several simulation estimators for a performance measure of a process having multiple regeneration sequences. We examine the setting of two regeneration sequences. We compare two existing estimators, the permuted estimator and the semi-regenerative estimator, and two new estimators, a type of U-statistic estimator and a type of V-statistic estimator. The last two estimators are obtained by resampling trajectories without and with replacement, respectively. The permuted estimator and the U-statistic estimator turn out to be equivalent, but the others are in general different. We show that when estimating the second moment of a cumulative cycle reward, the semi-regenerative and V-statistic estimators have non-negative bias, with the semi-regenerative bias being larger. The permuted estimator was previously shown to be unbiased. Although some of the estimators have different small-sample properties, they all satisfy central limit theorems with the same asymptotic variance constant.

1 INTRODUCTION

A regenerative process is a process that has an infinite sequence of stopping times, known as *regeneration points*, at which times the process probabilistically restarts. The evolution of the process between two successive regeneration points is called a *regenerative cycle*, and cycles are i.i.d. (e.g., see Shedler 1993). For example, for an irreducible, positive-recurrent Markov chain, the successive hitting times to a fixed state constitute a regeneration sequence. The standard regenerative method (Crane and Iglehart 1975) exploits the i.i.d. cycle structure in regenerative processes to construct asymptotically valid confidence intervals for a large class of performance measures.

Many regenerative processes possess several different regeneration sequences. For example, for a Markov chain, each fixed state yields a regeneration sequence, but there are many choices for the state to use. The standard regenerative method only uses one of these sequences, and in this paper, we examine several estimators that exploit more than one sequence. Throughout the paper, we consider estimating the second moment of the cumulative cycle reward of a discrete-time Markov chain on a discrete state space, but the ideas can also be applied to other performance measures and more general processes having multiple regeneration sequences. Also, we consider only two regeneration sequences.

Several methods to exploit multiple regeneration sequences have been proposed. Calvin and Nakayama (1998, 2000) developed *permuted regenerative estimators*, which can be described in the context of 2 regeneration sequences as follows. Run a simulation of a fixed number of cycles from one sequence, and based on the resulting sample path, calculate an estimate of the performance measure. Construct a new sample path from the original one by permuting the cycles from both sequences, and calculate an estimate based on the new path. Averaging over all permutations yields the permuted estimator. Calculation of this estimator does not require constructing all permutations since a simple closed-form representation for the estimator can be derived. Calvin and Nakayama (2000) show that the permuted estimator always has the same expectation as the standard regenerative estimator, so if the standard estimator is unbiased, so is the permuted estimator. Moreover, compared to the standard estimator, the permuted estimator always has no larger, and typically strictly smaller, variance.

Calvin, Glynn, and Nakayama (2002) developed another approach, the *semi-regenerative method*, which gets its name from its close connection to semi-regenerative processes (see Section 10.6 of Çinlar 1975). The idea is to fix a set A of states, and to break up the sample path into *trajectories* determined by successive entrances into A. Using the semi-regenerative structure, one derives an alternative representation of the performance measure in terms of expectations of random quantities defined over trajectories, and then constructs an estimator by replacing each expectation by its natural estimator. (Calvin, Glynn, and Nakayama 2002 also present a stratified semi-regenerative estimator in which trajectories are sampled in an i.i.d. fashion, but we will not consider that approach in this paper.) Other methods for simulating processes with multiple regeneration sequences include the *almost regenerative method* (Gunther and Wolff 1980) and *A-segments* (Zhang and Ho 1992).

In this paper, we present two new ways to construct estimators. Both approaches start with a sample path of a fixed number of cycles with respect to one regeneration sequence. The path is then broken into trajectories determined by visits into a set A, and the new estimators are obtained by resampling the trajectories and averaging over all possible resamples. One estimator considers resampling without replacement, which is a type of U-statistic, and the other resamples with replacement, which is a type of V-statistic. (See Serfling 1980, Chapter 5, for details on U- and V-statistics.)

We compare four estimators: the permuted, semi-regenerative, U-statistic, and V-statistic estimators. We show that the U-statistic and permuted estimators are equivalent. The other two estimators are different, and we derive the exact differences of all the estimators based on a sample path of a fixed number m of cycles from one regeneration sequence. It turns out that the semi-regenerative and V-statistic estimators are biased high for any m, with larger bias for the semi-regenerative estimator. Asymptotically, all four estimators are equivalent and satisfy central limit theorems with the same asymptotic variance constant.

The rest of the paper has the following organization. In Section 2 we describe the mathematical model and the performance measure considered. We derive the semi-regenerative estimator in Section 3, and we present the permuted estimator from Calvin and Nakayama (2000) in Section 4. Section 5 contains the resampled estimators, and we compare the four estimators in Section 6. We state some conclusions in Section 7. All proofs of theorems stated in this paper are given in Calvin and Nakayama (2002).

2 FRAMEWORK

For simplicity, we specialize the development to discrete-time Markov chains, but the results hold for more general processes with multiple regeneration sequences. Let $X = (X_j : j = 0, 1, 2, \ldots)$ be an irreducible positive-recurrent discrete-time Markov chain on a discrete state space $S = \{1, 2, 3, \ldots\}$. For any state $x \in S$, we can define a sequence of regeneration points $T(x) = (T_k(x) : k = 0, 1, 2, \ldots)$, where $T_0(x) = \inf\{j \geq 0 : X_j = x\}$ and $T_k(x) = \inf\{j > T_{k-1}(x) : X_j = x\}, k \geq 1$. We call the sample-path segment $(X_j : T_{k-1}(x) \leq j < T_k(x))$ the kth x-cycle.

Fix a state $w \in S$. For $x \in S$, define

$$\alpha(x) = E_x\left[\left(\sum_{k=0}^{\tau-1} f(X_k)\right)^2\right],$$

where $\tau = \inf\{k \geq 1 : X_k = w\}$ and $f : S \to \Re$ is a “reward’ function. Our goal is to estimate $\alpha(w)$.

We will compare the estimator of $\alpha(w)$ using the semi-regenerative method and the permuted estimator of Calvin and Nakayama (2000), and we will present two other estimators of $\alpha(w)$ based on resampling. Fix a set $A \subset S$ with $w \in A$. Define $T_0 = \inf\{j : X_j \in A\}$ and $T_k = \inf\{j \geq T_{k-1} + 1 : X_j \in A\}$ for $k \geq 1$.

We will carry out our analysis with $A = \{1, 2\}$ and $w = 1$. Also, we will assume that we simulate one sample path of a fixed number m of 1-cycles, and let the sample path be $\vec{X}_m = (X_j : j = 0, 1, \ldots, T_m(1))$, where $X_0 = 1$. Note that $\tau = T_1(1)$. Also, let $W = (W_k : k = 0, 1, 2, \ldots)$ with $W_k = X_{T_k}$ for $k \geq 0$, which is the embedded chain of X on visits to the set A. Let $M = m + |\{T_k(2), k \geq 0 : T_k(2) < T_m(1)\}|$, which is the number of returns to the set A up to time $T_m(1)$, so $T_M = T_m(1)$.

The standard estimator of $\alpha(1)$ based on the sample path $\vec{X}_m$ is

$$t_{\mathrm{STD}}(\vec{X}_m) = \frac{1}{m}\sum_{k=1}^{m} Y_k^2, \tag{1}$$

where

$$Y_k = \sum_{j=T_{k-1}(1)}^{T_k(1)-1} f(X_j)$$

for $k = 1, 2, \ldots, m$.

3 SEMI-REGENERATIVE ESTIMATOR

We now derive the semi-regenerative estimator for $\alpha(1)$. The key to applying the semi-regenerative method is deriving an alternative representation for $\alpha(1)$ based on returns to the set A. Let $T = \inf\{k \geq 1 : X_k \in A\}$. For $x \in A$, note that

$$\alpha(x) = E_x\left[\left(\sum_{k=0}^{T-1} f(X_k) + \sum_{y\in A} I(\tau > T, X_T = y)\sum_{k=T}^{\tau-1} f(X_k)\right)^2\right]$$

Expanding this formula for $\alpha(x)$, we see that it is composed of three terms:

$$\begin{aligned}\alpha(x) = E_x&\left[\left(\sum_{k=0}^{T-1} f(X_k)\right)^2\right]\\ +2\sum_{y\in A} E_x&\left[I(\tau > T, X_T = y)\left\{\sum_{k=T}^{\tau-1} f(X_k)\sum_{j=0}^{T-1} f(X_j)\right\}\right]\\ +\sum_{y\in A} E_x&\left[I(\tau > T, X_T = y)\left(\sum_{k=T}^{\tau-1} f(X_k)\right)^2\right]. \qquad (2)\end{aligned}$$

We now derive new expressions for the second and third terms in (2). Each summand in the second term in (2) satisfies

$$\begin{aligned}&E_x\left[I(\tau > T, X_T = y)\sum_{k=T}^{\tau-1} f(X_k)\sum_{j=0}^{T-1} f(X_j)\right]\\ &\quad = \; F(x, y)\, e(y),\end{aligned}$$

where

$$\begin{aligned}F(x, y) &= E_x\left[I(\tau > T, X_T = y)\sum_{k=0}^{T-1} f(X_k)\right],\\ e(y) &= E_y\left[\sum_{k=0}^{\tau-1} f(X_k)\right].\end{aligned}$$

Also

$$\begin{aligned}e(x) &= E_x\left[\sum_{k=0}^{T-1} f(X_k)\right]\\ &+ \sum_{y\in A} E_x\left[I(\tau > T, X_T = y)\right] e(y),\end{aligned}$$

so letting

$$q(x) = E_x\left[\sum_{k=0}^{T-1} f(X_k)\right],$$

and $G(x, y) = E_x\left[I(\tau > T, X_T = y)\right]$, $e = (e(x) : x \in A)$, $q = (q(x) : x \in A)$, and $G = (G(x, y) : x, y \in A)$, we see that $e = q + Ge$, or $e = (I - G)^{-1}q$. Thus, each summand in the second term in (2) satisfies

$$\begin{aligned}&E_x\left[I(\tau > T, X_T = y)\sum_{k=T}^{\tau-1} f(X_k)\sum_{j=0}^{T-1} X_j\right]\\ &\quad = \; F(x, y)\, ((I - G)^{-1}q)(y).\end{aligned}$$

For the summand in the third term in (2), we make the observation that

$$\begin{aligned}&E_x\left[I(\tau > T, X_T = y)\left(\sum_{k=T}^{\tau-1} f(X_k)\right)^2\right]\\ &\quad = \; G(x, y)\,\alpha(y).\end{aligned}$$

Now define

$$c(x) = E_x\left[\left(\sum_{k=0}^{T-1} f(X_k)\right)^2\right].$$

Let $\alpha = (\alpha(x) : x \in A)$, $c = (c(x) : x \in A)$, and $F = (F(x, y) : x, y \in A)$. Then, putting this all together, we get $\alpha = c + 2F(I - G)^{-1}q + G\alpha$, so

$$\alpha = (I - G)^{-1}(c + 2F(I - G)^{-1}q).$$

Now suppose $A = \{1, 2\}$ and $w = 1$. Note that $G(x, 1) = F(x, 1) = 0$ for $x \in A$. Let $J = (J(x, y) : x, y \in A)$ with $J = (I - G)^{-1}$, and note that

$$\begin{aligned}J &= \begin{pmatrix} 1 & -G(1,2) \\ 0 & 1 - G(2,2)\end{pmatrix}^{-1}\\ &= \begin{pmatrix} 1 & G(1,2)/(1 - G(2,2)) \\ 0 & 1/(1 - G(2,2))\end{pmatrix}.\end{aligned}$$

Thus,

$$\begin{aligned}\alpha &= \begin{pmatrix} 1 & J(1,2) \\ 0 & J(2,2)\end{pmatrix}\left[\begin{pmatrix} c(1) \\ c(2)\end{pmatrix}\right.\\ &+ \left.2\begin{pmatrix} 0 & F(1,2) \\ 0 & F(2,2)\end{pmatrix}\begin{pmatrix} 1 & J(1,2) \\ 0 & J(2,2)\end{pmatrix}\begin{pmatrix} q(1) \\ q(2)\end{pmatrix}\right]\\ &= \begin{pmatrix} c(1) + J(1,2)c(2) \\ J(2,2)c(2)\end{pmatrix}\\ &+ 2\begin{pmatrix} (F(1,2) + J(1,2)F(2,2))J(2,2)q(2) \\ J(2,2)^2F(2,2)q(2)\end{pmatrix},\end{aligned}$$

so

$$\begin{aligned}\alpha(1) &= c(1) + J(1,2)c(2)\\ &+ 2[F(1,2) + J(1,2)F(2,2)]J(2,2)q(2). \qquad (3)\end{aligned}$$

To derive an estimator based on (3) from the sample path $\vec{X}_m$ of m 1-cycles, we need estimates of the quantities on the right-hand side of (3). Recall that M was defined such that $T_M = T_m(1)$. For $x, y \in A$, we call a trajectory that begins in state x and ends in state y an (x, y)-trajectory, and define $h(x, y) = \sum_{k=0}^{M-1} I(W_k = x, W_{k+1} = y)$ as the number of such trajectories along the sample path $\vec{X}_m$. Let $H(x) =$

$\sum_{k=0}^{M-1} I(W_k = x)$, which is the number of trajectories in $\vec{X}_m$ that begin in state x. Note that $h(x,1)+h(x,2) = H(x)$ for $x \in A$. Also, $h(1,2) = h(2,1)$ since $\vec{X}_m$ is a path of a fixed number of 1-cycles, and $H(1) = m$.

For $x, y \in A$, let $T_1(x,y) = \inf\{T_k : k \geq 0,\ W_k = x,\ W_{k+1} = y\}$ and $T_i(x,y) = \inf\{T_k \geq T_{i-1}(x,y) + 1 : k \geq 0,\ W_k = x,\ W_{k+1} = y\}$ for $i \geq 2$. Also, let $T_1'(x,y) = \inf\{T_{k+1} : k \geq 0,\ W_k = x,\ W_{k+1} = y\}$ and $T_i'(x,y) = \inf\{T_{k+1} : k \geq 0,\ T_k \geq T_{i-1}(x,y) + 1,\ W_k = x,\ W_{k+1} = y\}$ for $i \geq 2$. Thus, for $k = 1, 2, \ldots, h(x,y)$, the kth trajectory starting in state $x \in A$ and ending in state $y \in A$ begins at time $T_k(x,y)$ and finishes at time $T_k'(x,y)$, and we define $Y_k(x,y)$ as the sum of the rewards along that trajectory; i.e.,

$$Y_k(x,y) = \sum_{j=T_k(x,y)}^{T_k'(x,y)-1} f(X_j).$$

Define $S_j(x,y) = \sum_{k=1}^{h(x,y)} Y_k(x,y)^j$.

Now for $x \in A$, observe that $(S_2(x,1)+S_2(x,2))/H(x)$ and $(S_1(x,1)+S_1(x,2))/H(x)$ are the natural estimators of $c(x)$ and $q(x)$, respectively. Also, we use $h(x,2)/H(x)$ and $S_1(x,2)/H(x)$ as estimators for $G(x,2)$ and $F(x,2)$, respectively. Our estimator of $J(1,2) = G(1,2)/(1 - G(2,2))$ is then

$$\begin{aligned} \hat{J}(1,2) &= \frac{h(1,2)}{H(1)}\left(1 - \frac{h(2,2)}{H(2)}\right)^{-1} \\ &= \frac{h(1,2)}{H(1)}\left(\frac{H(2)}{h(2,1)}\right) = \frac{H(2)}{H(1)} \end{aligned}$$

since $h(2,1) + h(2,2) = H(2)$ and $h(1,2) = h(2,1)$. Similarly, our estimator of $J(2,2) = 1/(1 - G(2,2))$ is $\hat{J}(2,2) = H(2)/h(1,2)$. Thus, the semi-regenerative estimator of $\alpha(1)$ based on the sample path $\vec{X}_m$ is

$$\begin{aligned} t_{\mathrm{SR}}(\vec{X}_m) = {} & \frac{1}{H(1)}(S_2(1,1)+S_2(1,2)) \\ & + \left(\frac{H(2)}{H(1)}\right)\frac{S_2(2,1)+S_2(2,2)}{H(2)} \\ & + 2\left[\frac{S_1(1,2)}{H(1)} + \left(\frac{H(2)}{H(1)}\right)\frac{S_1(2,2)}{H(2)}\right] \\ & \times \left(\frac{H(2)}{h(1,2)}\right)\frac{S_1(2,1)+S_1(2,2)}{H(2)} \\ = {} & \frac{1}{m}\Bigg(S_2(1,1)+S_2(1,2)+S_2(2,1)+S_2(2,2) \\ & + \frac{2}{h(1,2)}\big[S_1(1,2)S_1(2,1)+S_1(2,1)S_1(2,2) \\ & + S_1(1,2)S_1(2,2)+S_1(2,2)^2\big]\Bigg) \qquad (4) \end{aligned}$$

since $H(1) = m$. To simplify the comparison with other estimators, we define

$$\begin{aligned} Q &= \frac{1}{m}\Bigg(S_2(1,1)+S_2(1,2)+S_2(2,1)+S_2(2,2) \\ &+ \frac{2}{h(1,2)}\big[S_1(1,2)S_1(2,1)+S_1(2,1)S_1(2,2) \\ &+ S_1(1,2)S_1(2,2)\big]\Bigg), \qquad (5) \end{aligned}$$

so

$$t_{\mathrm{SR}}(\vec{X}_m) = Q + \frac{2}{mh(1,2)}S_1(2,2)^2. \qquad (6)$$

4 PERMUTED ESTIMATOR

The permuted estimator of $\alpha(1)$ based on the sample path $\vec{X}_m$ is obtained as follows. Recall we defined the standard estimator of $\alpha(1)$ as $t_{\mathrm{STD}}(\vec{X}_m)$ in (1). The permuted estimator of $\alpha(1)$ is obtained by permuting 2-cycles and 1-cycles to obtain a new sample path $\vec{X}_m'$, and averaging $t_{\mathrm{STD}}(\vec{X}_m')$ over all possible permuted paths $\vec{X}_m'$. Calvin and Nakayama (2000) showed that the permuted estimator is

$$\begin{aligned} & t_{\mathrm{P}}(\vec{X}_m) \\ & \equiv \frac{1}{m}\Bigg(S_2(1,1)+S_2(1,2)+S_2(2,1) \\ & + \frac{2}{h(1,2)}\Big[S_1(1,2)S_1(2,1) \\ & \quad + S_1(1,2)S_1(2,2)+S_1(2,1)S_1(2,2)\Big] \\ & + \frac{h(1,2)-1}{h(1,2)+1}S_2(2,2) + \frac{2}{h(1,2)+1}S_1(2,2)^2\Bigg) \\ & = Q + \frac{2}{m(h(1,2)+1)}\Big(S_1(2,2)^2 - S_2(2,2)\Big), \qquad (7) \end{aligned}$$

where Q is defined in (5).

5 RESAMPLED ESTIMATORS

We now present two estimators obtained via resampling the (x,y)-trajectories for $x, y \in A$. One estimator is based on resampling the (x,y)-trajectories without replacement (i.e., a permutation of trajectories), and the other resamples with replacement.

Fix a sample path $\vec{X}_m$. For $x, y \in A$, define $\Lambda(x,y)$ as the set of permutations of $(1, 2, \ldots, h(x,y))$, and let $\Lambda = \times_{x,y\in A}\Lambda(x,y)$. Let $\widehat{\Lambda}(x,y) = \{(i_1, i_2, \ldots, i_{h(x,y)}) : 1 \leq i_j \leq h(x,y) \text{ for } j = 1, 2, \ldots, h(x,y)\}$, which is the set of all $h(x,y)$-dimensional vectors in which the components are selected with replacement from

$\{1, 2, \ldots, h(x, y)\}$. Let $\widehat{\Lambda} = \times_{x,y\in A}\widehat{\Lambda}(x, y)$. Define $\Delta = \{(i_1, i_2, \ldots, i_{h(1,2)}) : i_1 = 1,\ i_{h(1,2)} \leq h(2, 2) + 1,\ i_j \leq i_{j+1}$ for $j = 1, 2, \ldots, h(1, 2) - 1\}$. Let $\Gamma = \Lambda \times \Delta$ and $\widehat{\Gamma} = \widehat{\Lambda} \times \Delta$, and let (K, D) denote a generic element of $\widehat{\Gamma}$, with $K = (K(x, y) : x, y \in A) \in \widehat{\Lambda}$, $K(x, y) = (K_i(x, y) : i = 1, 2, \ldots, h(x, y)) \in \widehat{\Lambda}(x, y)$, and $D = (D_1, D_2, \ldots, D_{h(1,2)}) \in \Delta$. Also, define $D_{h(1,2)+1} = h(2, 2) + 1$ and the function $g : \widehat{\Lambda} \times \Delta \to \Re$ as

$$
\begin{aligned}
g(K, D) &= \frac{1}{m}\Bigg[\sum_{i=1}^{h(1,1)} Y_{K_i(1,1)}(1, 1)^2 \\
&\quad + \sum_{i=1}^{h(1,2)} \Bigg(Y_{K_i(1,2)}(1, 2) + Y_{K_i(2,1)}(2, 1) \\
&\quad + \sum_{j=D_i}^{D_{i+1}-1} Y_{K_j(2,2)}(2, 2)\Bigg)^2\Bigg] \\
&= \frac{1}{m}\Bigg[\sum_{i=1}^{h(1,1)} Y_{K_i(1,1)}(1, 1)^2 \\
&\quad + \sum_{i=1}^{h(1,2)} \Bigg(Y_{K_i(1,2)}(1, 2)^2 + Y_{K_i(2,1)}(2, 1)^2 \\
&\quad + \left\{\sum_{j=D_i}^{D_{i+1}-1} Y_{K_j(2,2)}(2, 2)\right\}^2 \\
&\quad + 2Y_{K_i(1,2)}(1, 2)Y_{K_i(2,1)}(2, 1) \\
&\quad + 2\left\{Y_{K_i(1,2)}(1, 2) + Y_{K_i(2,1)}(2, 1)\right\} \\
&\quad \times \sum_{j=D_i}^{D_{i+1}-1} Y_{K_j(2,2)}(2, 2)\Bigg)\Bigg].
\end{aligned} \tag{8}
$$

Observe that the function g is conditional on the original sample path $\vec{X}_m$.

Note that $\Lambda(x, y)$ (resp., $\widehat{\Lambda}(x, y)$) is the set of all possible orderings of $h(x, y)$ (x, y)-trajectories, where the (x, y)-trajectories are chosen without (resp., with) replacement, and in (8), $K(x, y) \in \widehat{\Lambda}(x, y)$ is one such ordering of $h(x, y)$ (x, y)-trajectories. Also, note that Δ is the set of vectors specifying which $(2, 2)$-trajectories fall in between each pairing of $(1, 2)$-trajectory and $(2, 1)$-trajectory determined by $K(1, 2)$ and $K(2, 1)$; i.e., for $(D_1, \ldots, D_{h(1,2)}) \in \Delta$, "slots" D_j to $D_{j+1} - 1$ are the locations of the $(2, 2)$-trajectories in the ordered list $K(2, 2) \in \widehat{\Lambda}(2, 2)$ of $(2, 2)$-trajectories that fall between the $(1, 2)$-trajectory indexed by $K_j(1, 2)$ and the $(2, 1)$-trajectory indexed by $K_j(2, 1)$. Observe that if $D_j = D_{j+1}$, then there are no $(2, 2)$-trajectories that fall between the $(1, 2)$-trajectory indexed by $K_j(1, 2)$ and the $(2, 1)$-trajectory indexed by $K_j(2, 1)$.

Suppose we set $K^\circ(x, y) = (1, 2, \ldots, h(x, y))$ for all $x, y \in A$, and let $K^\circ = (K^\circ(x, y) : x, y \in A)$. Note that $K^\circ(x, y) \in \Lambda(x, y)$ for all $x, y \in A$, and $K^\circ \in \Lambda$. Define $D^\circ = (D_1^\circ, D_2^\circ, \ldots, D_{h(1,2)}^\circ)$ with $D_i^\circ = \min\{j : T_j(2, 2) > T_i(1, 2)\}$ for $i \geq 1$. Note that $D^\circ \in \Delta$. Then

$$g(K^\circ, D^\circ) = t_{\mathrm{STD}}(\vec{X}_m),$$

the standard estimator of $\alpha(1)$.

We now define the resampled estimators

$$t_U(\vec{X}_m) = \frac{1}{|\Gamma|}\sum_{(K,D)\in\Gamma} g(K, D) \tag{9}$$

and

$$t_V(\vec{X}_m) = \frac{1}{|\widehat{\Gamma}|}\sum_{(K,D)\in\widehat{\Gamma}} g(K, D), \tag{10}$$

where $t_U(\vec{X}_m)$ is the estimator based on resampling without replacement, and $t_V(\vec{X}_m)$ is the estimator with replacement. In addition to resampling trajectories, both estimators also average over all possible allocations of $h(2, 2)$ $(2, 2)$-trajectories to the 1-cycles containing a hit to state 2, which is accomplished in (9) and (10) by averaging over $D \in \Delta$. Note that $t_U(\vec{X}_m)$ is a type of U-statistic and $t_V(\vec{X}_m)$ is a type of V-statistic.

Figure 1 presents an example of a sample path $\vec{X}_m$ (top) and a path (middle) obtained from $\vec{X}_m$ by resampling without replacement the (x, y)-trajectories, for $x, y \in A$. State 1 corresponds to the horizontal axis, and state 2 corresponds to the dashed horizontal line. The original path $\vec{X}_m$, which is shown on top, has $m = 5$ 1-cycles. Also, the original path has $h(1, 1) = 2$ $(1, 1)$-trajectories, $h(1, 2) = 3$ $(1, 2)$-trajectories, $h(2, 1) = 3$ $(2, 1)$-trajectories, and $h(2, 2) = 3$ $(2, 2)$-trajectories. Moreover, $K_1(1, 1) = 1$, $K_2(1, 1) = 2$; $K_1(1, 2) = 1$, $K_2(1, 2) = 2$, $K_3(1, 2) = 3$; $K_1(2, 1) = 1$, $K_2(2, 1) = 2$, $K_3(2, 1) = 3$; $K_1(2, 2) = 1$, $K_2(2, 2) = 2$, $K_3(2, 2) = 3$; $D_1 = 1$, $D_2 = 2$, $D_3 = 4$, $D_4 = 4$. The resampled path shown on the bottom has $K_1(1, 1) = 2$, $K_2(1, 1) = 1$; $K_1(1, 2) = 3$, $K_2(1, 2) = 1$, $K_3(1, 2) = 2$; $K_1(2, 1) = 2$, $K_2(2, 1) = 3$, $K_3(2, 1) = 1$; $K_1(2, 2) = 3$, $K_2(2, 2) = 1$, $K_3(2, 2) = 2$; $D_1 = 1$, $D_2 = 1$, $D_3 = 3$, $D_4 = 4$.

The bottom path in Figure 1 is one possible path obtained from $\vec{X}_m$ by resampling with replacement the trajectories. The resampled path has $K_1(1, 1) = 2$, $K_2(1, 1) = 2$; $K_1(1, 2) = 3$, $K_2(1, 2) = 1$, $K_3(1, 2) = 3$; $K_1(2, 1) = 2$, $K_2(2, 1) = 3$, $K_3(2, 1) = 1$; $K_1(2, 2) = 3$, $K_2(2, 2) = 3$, $K_3(2, 2) = 2$; $D_1 = 1$, $D_2 = 1$, $D_3 = 3$, $D_4 = 4$. In the resampled path, $Y_2(1, 1)$ appears twice, and $Y_1(1, 1)$ does not appear. Also, for the $(1, 2)$-trajectories, $Y_1(1, 2)$ appears once, $Y_2(1, 2)$ does not appear, and $Y_3(1, 2)$ appears twice. For the $(2, 1)$-trajectories, each $Y_k(2, 1)$ appears once. For

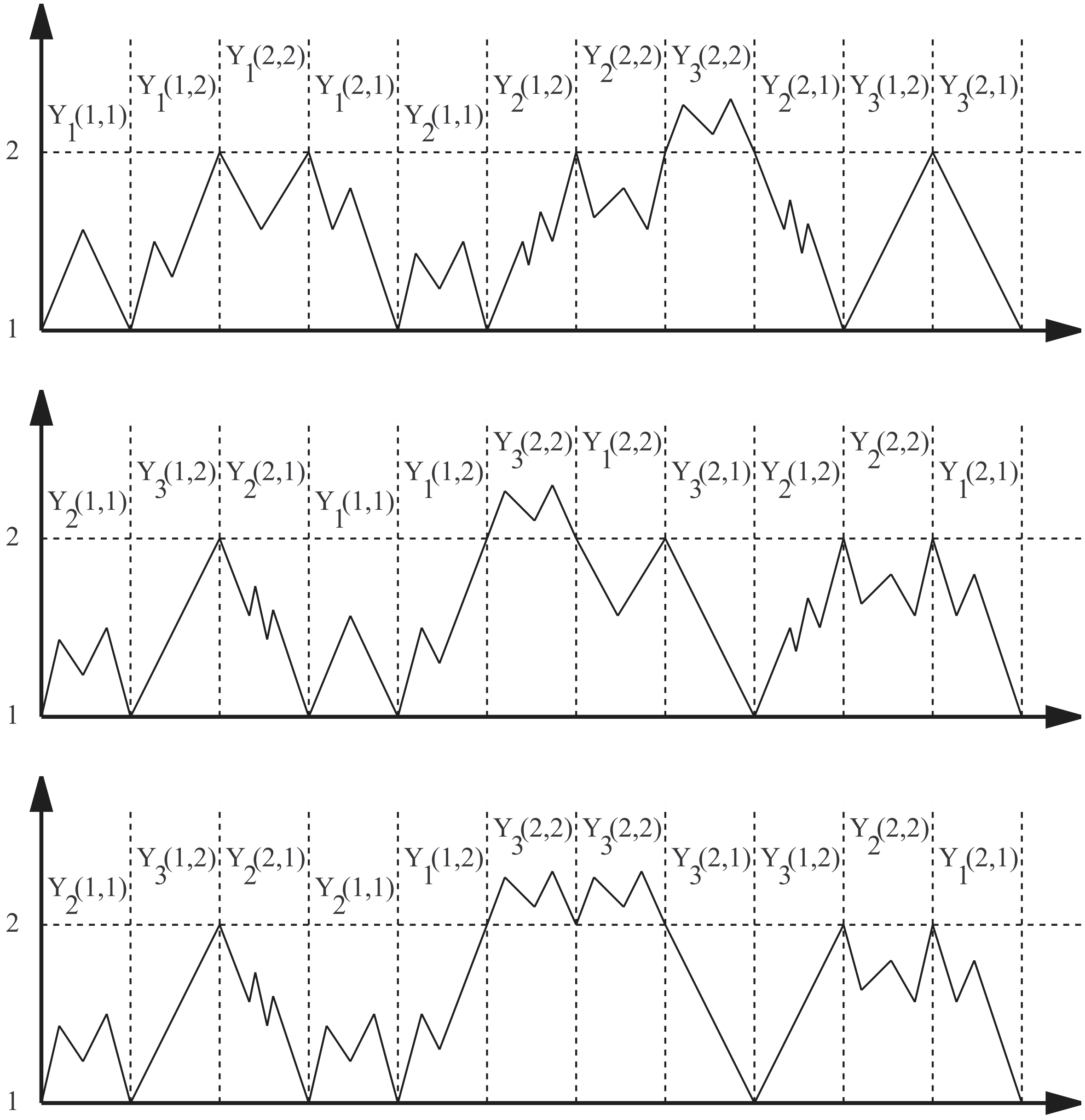

Figure 1: A Sample Path (top) and Paths Obtained by Resampling Trajectories Without Replacement (middle) and With Replacement (bottom)

the (2, 2)-trajectories, $Y_1(2,2)$ does not appear, $Y_2(2,2)$ appears once, and $Y_3(2,2)$ appears twice.

Theorem 1 *For all $\vec{X}_m$ and m, $t_U(\vec{X}_m) = t_P(\vec{X}_m)$ and*

$$t_V(\vec{X}_m) = Q + \frac{2(h(2,2)-1)}{m\,h(2,2)\,(h(1,2)+1)} S_1(2,2)^2, \qquad (11)$$

where Q is defined in (5).

Thus, since $t_U(\vec{X}_m) = t_{\mathrm{P}}(\vec{X}_m)$, we see that permuting trajectories and reallocating (2, 2)-trajectories to the 1-cycles leads to the same estimator as permuting cycles.

6 COMPARING THE ESTIMATORS

We start by comparing the semi-regenerative estimator $t_{\mathrm{SR}}(\vec{X}_m)$ in (6) and the permuted estimator $t_{\mathrm{P}}(\vec{X}_m)$ in (7). In the second term in the right-hand side of (7), the sum of the squares of the $Y_k(2,2)$ (i.e., $S_2(2,2)$) is subtracted from the square of the sum (i.e., $S_1(2,2)^2$), whereas the second term in the right-hand side of (6) only has the square of the sum. A similar situation occurs when comparing the V-statistic estimator $t_V(\vec{X}_m)$ in (11) and the permuted estimator $t_{\mathrm{P}}(\vec{X}_m)$.

The estimators of $\alpha(1)$ satisfy a complete ordering.

Theorem 2 *For all $\vec{X}_m$ and m,*

$$t_{SR}(\vec{X}_m) \geq t_V(\vec{X}_m) \geq t_U(\vec{X}_m) = t_P(\vec{X}_m), \qquad (12)$$

where

- *the first inequality in (12) is strict if and only if $S_1(2,2) \neq 0$;*
- *the second inequality in (12) is strict if and only if not all the $Y_k(2,2)$, $k = 1, 2, \ldots, h(2,2)$, are the same (i.e., there exists $i, j \in \{1, 2, \ldots, h(2,2)\}$ such that $Y_i(2,2) \neq Y_j(2,2)$);*
- *$t_{SR}(\vec{X}_m) > t_P(\vec{X}_m)$ if and only if $Y_k(2,2) \neq 0$ for some $k = 1, 2, \ldots, h(2,2)$.*

Moreover,

$$\begin{array}{rcccc} E[t_{SR}(\vec{X}_m)] & \geq & E[t_V(\vec{X}_m)] & \geq & E[t_U(\vec{X}_m)] \\ & = & E[t_P(\vec{X}_m)] & = & \alpha(1), \end{array} \qquad (13)$$

where

- *the first inequality in (13) is strict if and only if $P[S_1(2,2) = 0] < 1$;*
- *the second inequality in (13) is strict if and only if $Var[Y_1(2,2)] > 0$;*
- *$E[t_{SR}(\vec{X}_m)] > E[t_P(\vec{X}_m)]$ if and only if $P[Y_1(2,2) = 0] < 1$.*

Thus, $t_{SR}(\vec{X}_m)$ and $t_V(\vec{X}_m)$ are biased high.

We now compare some asymptotic properties of our estimators. Let $\xrightarrow{\mathcal{D}}$ denote convergence in distribution and $N(a, b)$ denote a normal distribution with mean a and variance b.

Theorem 3 *Assume $E[(\sum_{j=0}^{T_1(1)-1} |f(X_j)|)^4] < \infty$. Then for $t(\vec{X}_m)$ defined as either $t_P(\vec{X}_m)$, $t_{SR}(\vec{X}_m)$, or $t_V(\vec{X}_m)$, we have that $t(\vec{X}_m)$ converges with probability one to $\alpha(1)$ while a suitably standardized version of $t(\vec{X}_m)$ converges in distribution to a normal distribution with mean zero and variance σ_α^2, i.e.,*

$$t(\vec{X}_m) \to \alpha(1), \ \ a.s.,$$

and

$$\sqrt{m}\left[t(\vec{X}_m) - \alpha(1)\right] \xrightarrow{\mathcal{D}} N(0, \sigma_\alpha^2)$$

as $m \to \infty$, for some constant $\sigma_\alpha^2 > 0$, where σ_α^2 is the same for all three estimators and is given in Calvin and Nakayama (2000).

For all of the estimators of α, one can estimate the asymptotic variance constant σ_α^2 from one simulation run. Thus, one does not need to run multiple replications to obtain estimates of the asymptotic variance. See Calvin and Nakayama (2000) for details.

7 CONCLUSIONS

In this paper we showed that a wide spectrum of estimators can be obtained in simulations of processes having multiple regeneration sequences. We compared some of these estimators, two of which were proposed previously (the permuted and semi-regenerative estimators) and two are new (the U- and V-statistic estimators). We showed that the U-statistic and permuted estimators are equivalent. All of the other estimators are in general different, but they are asymptotically equivalent and they all satisfy the same central limit theorem. When estimating the second moment of a cycle reward, we showed that the estimators satisfy a complete ordering, with the semi-regenerative estimator being the largest, the V-statistic estimator being next largest, and the U-statistic and permuted estimators being the smallest. More work is needed to analyze the behavior of these estimators for other performance measures and to determine which is the "best" (perhaps in terms of smallest mean squared error) for a specific measure in the small-sample context.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grants DMI-9624469 and DMI-9900117.

REFERENCES

Calvin, J. M. and M. K. Nakayama. 1998. Using permutations in regenerative simulations to reduce variance. *ACM Transactions on Modeling and Computer Simulation* 8: 153–193.

Calvin, J. M. and M. K. Nakayama. 2000. Simulation of processes with multiple regeneration sequences. *Probability in the Engineering and Informational Sciences*, 14: 179–201.

Calvin, J. M. and M. K. Nakayama. 2002. Resampled regenerative estimators. Forthcoming.

Calvin, J. M., P. W. Glynn and M. K. Nakayama. 2002. The semi-regenerative method of simulation output analysis. Forthcoming.

Çinlar, E. 1975. *Introduction to Stochastic Processes*. Englewood Cliffs, N.J.: Prentice-Hall.

Crane, M. and D. L. Iglehart. 1975. Simulating stable stochastic systems, III: Regenerative processes and discrete-event simulations. *Operations Research* 23: 33–45.

Gunther, F. L. and R. W. Wolff. 1980. The almost regenerative method for stochastic system simulations. *Operations Research* 28: 375–386.

Serfling, R.J. 1980. *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Shedler, G. S. 1993. *Regenerative Stochastic Simulation*. San Diego: Academic Press.

Zhang, B. and Y.-C. Ho. 1992. Improvements in the likelihood ratio method for steady-state sensitivity analysis and simulation. *Performance Analysis* 15: 177–195

AUTHOR BIOGRAPHIES

JAMES M. CALVIN is an associate professor in the Department of Computer Science at the New Jersey Institute of Technology. He received a Ph.D. in operations research from Stanford University and is an associate editor for *ACM Transactions on Modeling and Computer Simulation*. Besides simulation output analysis, his research interests include global optimization and probabilistic analysis of algorithms.

MARVIN K. NAKAYAMA is an associate professor in the Department of Computer Science at the New Jersey Institute of Technology. He received a Ph.D. in operations research from Stanford University. He won second prize in the 1992 George E. Nicholson Student Paper Competition sponsored by INFORMS and is a recipient of a CAREER Award from the National Science Foundation. He is the area editor for the Stochastic Modeling Area of *ACM Transactions on Modeling and Computer Simulation* and an associate editor for *Informs Journal on Computing*. His research interests include applied probability, statistics, simulation and modeling.