965-78

Request Date: 02-SEP-2010
Expiration Date: 31-OCT-2010
ILL Number:
ILL Number: 4080851
Call Number: TA329 .A649
Format: Article Printed
Title: Transactions of the ... Army Conference on Applied Mathematics and Computing
Article Author: Peter Glynn
Article Title: A Non-Rectangular Sampling Plan for Estimating Steady-State Means
Volume/Issue: 6TH /
Part Pub. Date: / 1988
Pages: 965-978
Publication Date: 1988
Borrower: ST2

TGQ or OCLC #:
TGQ or OCLC #: 66202
ID: ST2
ISBN/ISSN:

Address: Stanford University
Green
Interlibrary Services, Green Library room 161
Stanford, CA
94305-6004

Patron Name: Rhee, Changhan
Patron e-mail:
Service Level:
Delivery Method:
Request Notes:
Need By:
Verification Source:
Supplier Reference:
Owned By: NRLF

Service Type: Copy non returnable
Max Cost: USD15
Payment Type: IFM

Copyright Info: US:CCL
Requester Symbol: OCLC:ST2
Return To: 1301 So. 46th St.
RFS-400
Richmond, CA
94804-4698

C 3 089 550-551

stf.072

AD-A207 252
PART 2 OF 2

ARO Report 89—1

TRANSACTIONS OF THE SIXTH ARMY CONFERENCE ON APPLIED MATHEMATICS AND COMPUTING

Sponsored by

The Army Mathematics Steering Committee

on behalf on

THE CHIEF OF RESEARCH, DEVELOPMENT

AND ACQUISITION

25 055

A NON-RECTANGULAR SAMPLING PLAN FOR ESTIMATING STEADY-STATE MEANS

Peter W. Glynn
Department of Operations Research
Stanford University
Stanford, CA 94305

Abstract

The method of multiple replicates is frequently used by simulators to estimate the steady-state mean of a stochastic simulation. One important advantage of this approach is that it is easily adapted to a parallel computer. Unfortunately, the method of multiple replicates is quite sensitive to contamination by "initial bias." In this paper, a new type of sampling plan is described. It retains the replication flavor, yet attenuates the bias problem. It is shown that the new method reduces mean square error relative to conventional multiple replicates for problems in which the "initial transient" decays slowly.

Keywords: Simulation, replication, mean square error, parallel computation.

Introduction

Let $Y = (Y(n) : n \geq 0)$ be a real-valued stochastic sequence corresponding to the output of a stochastic simulation. We assume that Y is ergodic, in the sense that there exists a finite (deterministic) constant r such that

$$\frac{1}{n}\sum_{i=0}^{n-1} Y(i) \Rightarrow r$$

as $n \to \infty$. The steady-state simulation problem concerns the question of estimating the parameter r efficiently, and providing confidence intervals for r.

Basically, two alternative approaches for dealing with this problem have been studied in the literature. One approach is known as the method of multiple replicates. The idea here is to generate m independent replicates of the process Y. Each replicate is simulated for t time units. The advantage of this method is that it gives rise to independent observations; this significantly simplifies the problem of producing confidence intervals for r. Furthermore, given access to a parallel computing environment, one can assign each independent replicate to a different processor. Thus, the method of multiple replicates is well suited to parallel computation.

A disadvantage of this approach is that each of the m independent replicates is contaminated by initial bias. This initial bias arises from the fact that each of the m replicates is initiated with an initial condition that is atypical of the steady-state of the system. If we view the first s time units of each replicate as representing an "initial transient" for the system, this analysis suggests that ms time units of the total time simulated are contaminated by initial bias. If m is large, we find that the method of multiple replicates devotes a significant amount of computation to generation of highly biased observations. This is, of course, undesirable.

In response to this, we can consider sampling plans in which only one observation of Y is generated. Such a strategy is known in the literature as a single replication method. Here, only the first s time units of the simulation are significantly biased, and there is no magnification effect by the parameter m. On the other hand, construction of confidence intervals for r is now complicated by the fact that all the observations collected are autocorrelated. Furthermore, it is now a non-trivial task to make an assignment of parallel processors that will significantly speed up the simulation.

Note that the method of multiple replicates involves factoring a computer time budget T into m replicates, each of length $t = T/m$. If we view the data of the i'th replicate as being assigned to the i'th row of a matrix, we obtain a rectangular $m \times t$ matrix which summarizes the data generated by the simulation. Consequently, we refer to the method of multiple replicates as a rectangular sampling plan for estimating steady-state means (see Figure 1). Of course, a single replicate method is the special case of a rectangular scheme in which the data corresponds to a $1 \times T$ row vector.

In this paper, we consider these rectangular methods in greater detail. We also propose and analyze a new non-rectangular sampling scheme, which attempts to offer an advantageous compromise between the methods of single and multiple replicates.

The organization of this paper is as follows. Section 2 provides reasonably complete mean square error analysis of conventional rectangular sampling plans. In Section 3, the non-rectangular plan is introduced and studied. Section 4 offers some conclusions.

2. Rectangular Sampling Plans

We start by describing the traditional method of replication for solving the steady-state simulation problem. To simplify the discussion that follows, we will assume that in x units of computer time, precisely x time units of the process Y can be simulated. Thus, given a total computer time budget of size T, we can implement a rectangular sampling plan in the following way:

1.) Choose the number m of independent replicates. (If $m = 1$, this is a single replication method.)

2.) Choose the (deletion) parameter s, from the interval $[0, T/m]$. (The first s time units of each replication will be deleted from the set of observations.)

3.) Generate m independent copies $Y_1, Y_2, \ldots, Y_m$ of the process Y. Each copy is simulated over the interval $[0, T/m]$.

4.) Set $t = \lfloor T/m \rfloor$ and compute the estimator

$$\bar{Y}(m, s, T) = \frac{1}{m(t-s)} \sum_{i=1}^{m} \sum_{j=s+1}^{t} Y_i(j).$$

We will now consider the mean square error (MSE) of the estimator $\bar{Y}(m, s, T)$. The MSE criterion is often viewed as the most important quantitative measure of the quality of an estimator. We start with the well known MSE decomposition formula

$$\text{MSE}(\bar{Y}(m, s, T)) = \text{var}\, \bar{Y}(m, s, T) + (\text{bias}\, \bar{Y}(m, s, T))^2. \tag{2.1}$$

By using the independence of the replicates, we observe that

$$\text{var}\, \bar{Y}(m, s, T) = \frac{1}{m} \text{var} \frac{1}{t-s} \sum_{j=s+1}^{t} Y(j), \tag{2.2}$$

$$\text{bias}\, \bar{Y}(m, s, T) = \frac{1}{t-s} \sum_{j=s+1}^{t} EY(j) - r. \tag{2.3}$$

A Rectangular Sampling Plan

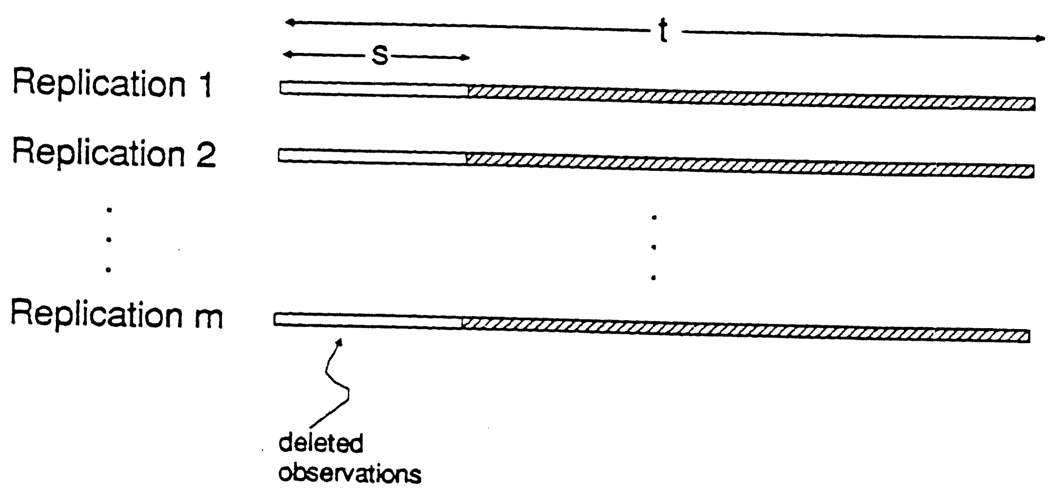

Figure 1

The Non-rectangular Sampling Plan

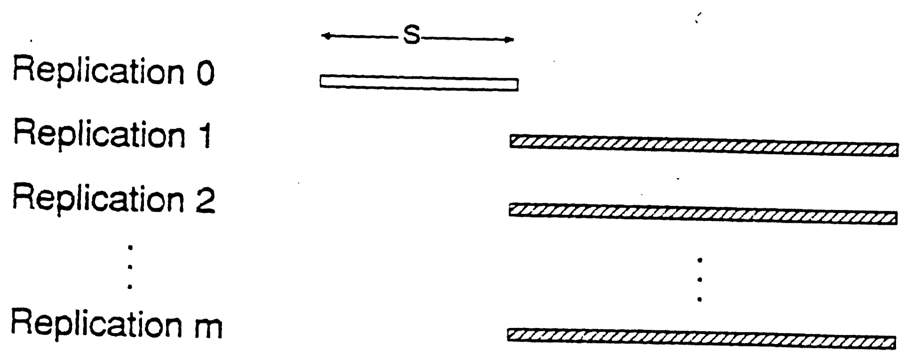

Figure 2

In order to analyze the terms appearing on the right-hand sides of (2.2) and (2.3), we will assume that $Y(n)$ can be expressed as a real-valued functional of a time-homogeneous Markov chain $X(n)$, so that $Y(n) = f(X(n))$ for some real-valued f defined on the state space S of X. The set S may be discrete or continuous. Continuous state space is particularly convenient in analysis of discrete-event simulations. The generalized semi-Markov process (GSMP) view of discrete-event systems shows that very general discrete-event simulations may be expressed in the form $Y(n) = f(X(n))$ with X Markov, provided that we permit continuous state space.

For $x \epsilon S, u \geq 1$, let $v(x, u)$ be the conditional variance defined by

$$v(x,u) = E\left\{\left(\frac{1}{u}\sum_{j=0}^{u-1} Y(j)\right)^2 |X(0) = x\right\} - \left(E\left\{\frac{1}{u}\sum_{j=0}^{u-1} Y(j)|X(0) = x\right\}\right)^2.$$

Similarly, let $b(x, u)$ be the conditional bias given by

$$b(x,u) = E\left\{\frac{1}{u}\sum_{j=0}^{u-1} Y(j)|X(0) = x\right\} - r.$$

Let $\mu(\cdot) = P\{X(0)\epsilon\cdot\}$ be the initial distribution of X. The Markov property permits us to re-express (2.3) as

$$\text{bias}\ \bar{Y}(m, s, T) = E_\mu b(X(s+1), t-s), \tag{2.4}$$

where $E_\mu(\cdot)$ denotes the expectation operator conditional on $X(0)$ having distribution μ.

To obtain a similar expression for the variance term (2.2) requires more care. We first apply the well known variance decomposition formula

$$\begin{aligned}\text{var}\frac{1}{t-s}\sum_{j=s+1}^{t} Y(j) = &\ E\,\text{var}\left\{\frac{1}{t-s}\sum_{j=s+1}^{t} Y(j)|X(s+1)\right\} \\ &+ \text{var}\,E\left\{\frac{1}{t-s}\sum_{j=s+1}^{t} Y(j)|X(s+1)\right\}.\end{aligned} \tag{2.5}$$

Clearly, we have

$$\text{var}\left\{\frac{1}{t-s}\sum_{j=s+1}^{t} Y(j)|X(s+1)\right\} = v(X(s+1), t-s),$$

$$E\left\{\frac{1}{t-s}\sum_{j=s+1}^{t} Y(y)|X(s+1)\right\} = b(X(s+1), t-s).$$

Plugging these expressions into (2.5) yields

$$\text{var}\frac{1}{t-s}\sum_{j=s+1}^{t} Y(y) = E_\mu v(X(s+1), t-s) + \text{var}_\mu b(X(s+1), t-s), \tag{2.6}$$

where $\mathrm{var}_\mu(\cdot)$ denotes the variance operator conditional on $X(0)$ having distribution μ.

Suppose that X is a positive recurrent Markov chain possessing a unique invariant probability distribution π. A large class of such chains has the property that under suitable regularity conditions,

$$\sup_{h \in \mathcal{H}} |E_\mu h(X(s)) - E_\pi h(X(0))| = 0(e^{-\alpha s})$$

for some $\alpha > 0$, where $\mathcal{H}$ is some appropriately defined family of real-valued functions $h : S \to \mathbb{R}$. (See NUMMELIN (1984), p. 120, for an example of such a theorem.) Assuming that the functions $v(\cdot, u)$, $b(\cdot, u)$, $b^2(\cdot, u) \varepsilon \mathcal{H}$ for all $u \geq 1$, we obtain the relations

$$E_\mu v(X(s+1), t-s) = E_\pi V(X(0), t-s) + O(e^{-\alpha s}), \tag{2.7}$$

$$E_\mu b(X(s+1), t-s) = E_\pi b(X(0), t-s) + O(e^{-\alpha s}), \tag{2.8}$$

$$E_\mu b^2(X(s+1), t-s) = E_\pi b^2(X(0), t-s) + O(e^{-\alpha s}), \tag{2.9}$$

where the constants implicit in each of the "big Oh" terms are independent of t.

Furthermore, for such a recurrent Markov chain, it is typically the case that the steady-state mean r can be expressed in the form $r = E_\pi f(X(0))$. As a consequence of the stationarity of X under initial distribution π, it is evident that $E_\pi Y(n) = r$ for $n \geq 0$ and hence $E_\pi b(X(0), t-s) = 0$. Thus, (2.8) can be simplified to

$$E_\mu b(X(s+1), t-s) = O(e^{-\alpha s}). \tag{2.10}$$

Combining (2.9) and (2.10), we obtain

$$\mathrm{var}_\mu b(X(s+1), t-s) = E_\pi b^2(X(0), t-s) + O(e^{-\alpha s}). \tag{2.11}$$

(Again, the constants implicit in (2.10) and (2.11) are independent of t.)

Combining (2.6), (2.7), and (2.11), we obtain the expression

$$\mathrm{var}\, \frac{1}{t-s} \sum_{j=s+1}^{t} Y(j) = E_\pi v(X(0), t-s) + E_\pi b^2(X(0), t-s) + O(e^{-\alpha s}).$$

Repeating the variance decomposition (2.6) under $\mathrm{var}_\pi(\cdot)$, we find that

$$\mathrm{var}_\pi \frac{1}{t-s} \sum_{j=s+1}^{t} Y(j) = E_\pi v(X(0), t-s) + E_\pi b^2(X(0), t-s)$$

and hence

$$\mathrm{var}\, \frac{1}{t-s} \sum_{j=s+1}^{t} Y(j) = \mathrm{var}_\pi \frac{1}{t-s} \sum_{j=s+1}^{t} Y(j) + O(e^{-\alpha s}). \tag{2.12}$$

To simplify (2.12), we again use the fact that X is stationary under initial distribution π. Set $Y_c(n) = Y(n) - r$,

$$\sigma^2 = E_\pi Y_c(0)^2 + 2\sum_{k=1}^{\infty} E_\pi Y_c(0) Y_c(k)$$

$$\eta = 2\sum_{k=1}^{\infty} k E_\pi Y_c(0) Y_c(k).$$

Under appropriate summability hypotheses (see, for example, p. 172 of BILLINGSLEY (1968)), we can use the stationarity to write

$$\text{var}_\pi \frac{1}{t-s}\sum_{j=s+1}^{t} Y(j) = \frac{\sigma^2}{t-s} - \frac{\eta}{(t-s)^2} - \frac{2}{t-s}\sum_{k=t-s}^{\infty}\left(1 - \frac{k}{t-s}\right) E_\pi Y_c(0) Y_c(k). \tag{2.13}$$

Note that

$$|E_\pi Y_c(0) Y_c(k)| \leq \int_S |f(x)| \cdot |E_x Y_c(k)| \cdot \pi(dx), \tag{2.14}$$

where $E_x(\cdot)$ is the expectation operator conditional on $X(0) = x$. We now observe that $E_x Y(k) = E_x f(X(k)) - E_\pi f(X(0))$. Appropriate regularity hypotheses on X permit us to assert that

$$\sup_{x \in S} |E_x f(X(k)) - E_\pi f(X(0))| = O(e^{-\beta k}) \tag{2.15}$$

for some $\beta > 0$. (See p. 122 of NUMMELIN (1984) for a typical such result.) Substituting this relation in (2.14) yields

$$E_\pi Y_c(0) Y_c(k) = O(e^{-\beta k}).$$

We may therefore conclude that

$$\sum_{k=u}^{\infty}\left(1 - \frac{k}{u}\right) E_\pi Y_c(0) Y_c(k) = O(e^{-\beta' u}) \tag{2.16}$$

for $0 < \beta' < \beta$. Substitution of (2.16) into (2.13) shows that

$$\text{var}_\pi \frac{1}{t-s}\sum_{j=s+1}^{t} Y(j) = \frac{\sigma^2}{t-s} - \frac{\eta}{(t-s)^2} + O(e^{-\beta'(t-s)}). \tag{2.17}$$

Combining (2.1), (2.2), (2.4), (2.10), (2.12), and (2.17), we obtain the important relationship

$$\text{MSE}(\bar{Y}(m, s, T)) = \frac{1}{m}\left(\frac{\sigma^2}{t-s} - \frac{\eta}{(t-s)^2}\right) + O(e^{-\alpha s}) + \frac{1}{m} O(e^{-\beta'(t-s)}), \tag{2.18}$$

where the implicit constants appearing in the "big OH" terms are independent of m, s, and T.

To gain further insight into (2.18), we consider the typical situation, in which the deletion point s is small relative to the length t of each replicate. Furthermore, in order to simplify the discussion, we assume that $mt = T$ (exactly). Then,

$$\frac{1}{m}\frac{\sigma^2}{t-s} = \frac{\sigma^2}{T} + \frac{\sigma^2 ms}{T^2} + \frac{1}{T}O\left(\frac{s^2m^2}{T^2}\right), \quad \text{and} \tag{2.19}$$

$$\frac{1}{m}\frac{\eta}{(t-s)^2} = \frac{m\eta}{T^2} + \frac{m}{T^2}O\left(\frac{ms}{T}\right). \tag{2.20}$$

Combining (2.18) through (2.20), we obtain the approximation

$$\text{MSE}(\bar{Y}(m,s,T)) \approx \frac{\sigma^2}{T} + \frac{\sigma^2 ms}{T^2} - \frac{m\eta}{T^2}. \tag{2.21}$$

Viewing s and m as design parameters for the simulation, we see that (2.21) suggests that the deletion parameter s should be small. On the other hand, if s is chosen too small, difficulties can arise in the "big Oh" terms appearing in (2.18). This recommendation corresponds to intuition.

As for the number of replications m, m should be chosen small (for example, a single replicate method should be considered) whenever $\sigma^2 s > \eta$. For reasonable values of s, this inequality will typically be valid. Thus, mean square error favors using a small number of replicates. This differs from the conclusion reached by KELTON (1986) in his analysis of "replication splitting" schemes for simulation of autoregressive sequences. The arguments there show that using a large number of replicates can reduce the variance of the steady-state estimator when the autoregressive sequence is positively correlated (i.e. $\eta > 0$). In our current setting, we judge our estimators via mean square error (as opposed to variance). Since our error criterion explicitly considers the loss in estimator efficiency due to bias (variance does not measure bias), it is not surprising that our conclusions differ. Of course, if s is small (i.e. bias is not a major problem), (2.21) supports using a large number of replicates when $\eta > 0$,

To illustrate the above points, we calculate the mean square error of $\bar{Y}(m,s,T)$ when $m = T^p (0 \le p < 1)$ and $s = T^q (0 < q < 1-p)$, in which case $t = T^r$, where $r = 1-p$. We find that

$$\text{MSE}(\bar{Y}(m,s,T)) = \frac{\sigma^2}{T} + \frac{\sigma^2}{T^{2-p-q}} - \frac{\eta}{T^{2-p}} + O(T^{2p+2q-3}). \tag{2.22}$$

Assuming that $p+q < 1/2$, (so that the "big oh" term is small) we find that relation (2.17) confirms the previous discussion. Both p and q should be chosen small, in accordance with our previous recommendations.

3. A Non-Rectangular Sampling Plan

The idea behind the sampling plan to be described in this section is that we try to avoid expending a significant fraction of the computer time budget on generation of highly biased observations. As discussed in the Introduction, the initial bias problem is of particular concern when the method of multiple replicates is used, since the amount of contaminated data is proportional to the number of replicates. On the other hand, the method of multiple replicates enjoys several significant advantages: ease of construction of confidence intervals and development of parallel simulation schemes. Our goal here is to develop a method that has a multiple replicate flavor and yet avoids the initial bias difficulties that are associated with conventional multiple replicate methods.

As in Section 2, we assume that the output sequence Y takes the form $Y(n) = f(X(n))$ for some time-homogeneous Markov chain X, and real-valued function f. The following algorithm employs one simulation of length s to generate an initial condition which is reasonably typical of the steady-state. This initial condition is then used to generate m conditionally independent replicates (each of length t) from the output sequence Y. Thus, the effort to generate a "good" initial condition is amortized over the m replicates. In terms of observations generated, this sampling plan is non-rectangular (see Figure 2).

The non-rectangular sampling plan can be summarized as follows.

1.) Given the computer time budget T, choose the number m of (conditionally independent) replicates, and the deletion parameter s $(0 \leq s \leq T)$.
2.) Generate one copy Y_0 of the sequence Y to time s.
3.) Using the initial condition $X_0(s)$ (X_0 is the Markov chain corresponding to Y_0), generate m copies $Y_1, \ldots, Y_m$ of Y to time $t-1$, where $t = \lceil (T-s)/m \rceil$.
4.) Compute the estimator

$$\tilde{Y}(m,s,T) = \frac{1}{mt} \sum_{i=1}^{m} \sum_{j=0}^{t-1} Y_i(j)$$

We now turn to computing the mean square error of $\tilde{Y}(m,s,T)$. As in Section 2,

$$\text{MSE}(\tilde{Y}(m,s,T)) = \text{var}\, \tilde{Y}(m,s,T) + (\text{bias}\, \tilde{Y}(m,s,T))^2. \tag{3.1}$$

Using the fact that $Y_i(\cdot) \stackrel{D}{=} Y(\cdot + s)$ ($\stackrel{D}{=}$ denotes equality in distribution), we find that

$$\text{bias}\, \tilde{Y}(m,s,T) = E_\mu b(X(s), t).$$

From (2.8), it therefore follows that

$$\text{bias}\, \tilde{Y}(m,s,T) = O(e^{-\alpha s}). \tag{3.2}$$

To handle the variance term appearing on the right-hand side of (3.1), we again use the variance decomposition method:

$$\text{var}\, \tilde{Y}(m,s,T) = \text{var}\, E\{\tilde{Y}(m,s,T) | X_0(s)\} + E\, \text{var}\{\tilde{Y}(m,s,T) | X_0(s)\}. \tag{3.3}$$

It is easily seen (use the fact that $Y_1, \ldots, Y_m$ are independent and identically distributed, conditional on $X_0(s)$) that

$$E\{\tilde{Y}(m, s, T) | X_0(s)\} = b(X_0(s), t) \quad \text{a.s.,} \tag{3.4}$$

$$\text{var}\{\tilde{Y}(m, s, T) | X_0(s)\} = \frac{1}{m} v(X_0(s), t) \quad \text{a.s.} \tag{3.5}$$

Combining (3.3) through (3.5), we get

$$\text{var}\, \tilde{Y}(m, s, T) = \frac{1}{m} E_\mu v(X(s), t) + \text{var}_\mu b(X(s), t). \tag{3.6}$$

As in Section 2, we obtain

$$\text{var}\, \tilde{Y}(m, s, T) = \frac{1}{m} E_\pi v(X(0), t) + E_\pi b^2(X(0), t) + O(e^{-\alpha s}) \tag{3.7}$$

(use (2.7), (2.8), and (2.9)). Recall that

$$\text{var}_\pi \frac{1}{t} \sum_{j=0}^{t-1} Y(j) = E_\pi v(X(0), t) + E_\pi b^2(X(0), t).$$

(see Section 2). Plugging into (3.7), we get

$$\text{var}\, \tilde{Y}(m, s, T) = \frac{1}{m} \text{var}_\pi \frac{1}{t} \sum_{j=0}^{t-1} Y(j) + \left(\frac{m-1}{m}\right) E_\pi b^2(X(0), t). \tag{3.8}$$

The first term on the right-hand side of (3.8) was analyzed in (2.17). For the second term, note that

$$b(x, t) = \frac{1}{t} b(x) - \frac{1}{t} \sum_{k=t}^{\infty} (E_x Y(k) - r),$$

where

$$b(x) = \sum_{k=0}^{\infty} (E_x Y(k) - r).$$

From (2.15), it is evident that

$$\sup_{x \in S} |b(x, t) - \frac{1}{t} b(x)| = O(e^{-\beta t}). \tag{3.9}$$

Consequently, we obtain the inequality

$$b(X(0), t) \leq \frac{1}{t} b(X(0)) + O(e^{-\beta t}). \tag{3.10}$$

Since $E_\pi Y(k) = r$, the expectations $E_\pi b(X(0), t)$ and $E_\pi b(X(0))$ both vanish. From (3.10), we therefore get

$$E_\pi b^2(X(0), t) \leq \frac{1}{t^2} E_\pi b^2(X(0)) + O(e^{-\beta t}) E_\pi |b(X(0))| + O(e^{-2\beta t}).$$

A similarly derived lower bound yields the formula

$$E_\pi b^2(X(0),t) = \frac{1}{t^2}E_\pi b^2(X(0)) + O(e^{-\beta t}). \tag{3.11}$$

Let $b = E_\pi b^2(X(0))$. To simplify the following discussion, assume $t = (T-s)/m$ (exactly). Combining (2.13), (3.8), and (3.11), we obtain the important relationship

$$\mathrm{MSE}(\tilde{Y}(m,s,T)) = \frac{1}{m}\left(\frac{\sigma^2}{t} - \frac{\eta}{t^2}\right) + \left(\frac{m-1}{m}\right)\frac{b}{t^2} + O(e^{-\gamma t}) + O(e^{-\alpha s}), \tag{3.12}$$

where $\gamma = \min(\beta, \beta')$ and the (implicit) constants in the "big oh" terms are independent of m, s, and T. Expressing t in terms of m, s, and T, we get

$$\frac{\sigma^2}{mt} = \frac{\sigma^2}{T} + \frac{\sigma^2 s}{T^2} + \frac{1}{T}O\left(\frac{s^2}{T^2}\right), \tag{3.13}$$

$$\frac{\eta}{mt^2} = \frac{m\eta}{T^2} + \frac{m}{T^2}O\left(\frac{s}{T}\right), \quad \text{and} \tag{3.14}$$

$$\left(\frac{m-1}{m}\right)\frac{b}{t^2} = \frac{(m-1)mb}{T^2} + \frac{m(m-1)}{T^2}O\left(\frac{s}{T}\right), \tag{3.15}$$

assuming that s is small relative to T. Combining (3.12) through (3.15), we obtain the approximation

$$\mathrm{MSE}(\tilde{Y}(m,s,T)) \approx \frac{\sigma^2}{T} + \frac{\sigma^2 s}{T^2} - \frac{m\eta}{T^2} + \frac{m(n-1)b}{T^2}. \tag{3.16}$$

We now compare the mean square error of our non-rectangular sampling plan with that of a rectangular plan having the same computer time budget T, number of replications m, and deletion parameter s. Comparing (3.16) to (2.21), we see that $\mathrm{MSE}(\tilde{Y}(m,s,T)) \leq \mathrm{MSE}(\bar{Y}(m,s,T))$ when

$$\sigma^2 ms \geq \sigma^2 s + b(m^2 - m).$$

We shall shortly show that $b \geq \sigma^2$. Thus, $\tilde{Y}(m,s,T)$ beats $\bar{Y}(m,s,T)$ when $sm \geq s + m^2$. This will typically occur when s is large relative to m. Thus, we can expect $\tilde{Y}(m,s,T)$ to have smaller MSE than $\bar{Y}(m,s,T)$ whenever s must be chosen relatively large, in order to remove initial bias.

We can illustrate this point when $m = T^p$ $(0 \leq p < 1)$ and $s = T^q$ $(0 < q < 1)$. Then, if $p + q < 1/2$,

$$\mathrm{MSE}(\tilde{Y}(m,s,T)) = \frac{\sigma^2}{T} + \frac{\sigma^2}{T^{2-q}} - \frac{\eta}{T^{2-p}} + \frac{b}{T^{2-2p}} + O\left(\frac{1}{T^2}\right). \tag{3.17}$$

Comparing (3.17) to (2.22), we find that $\mathrm{MSE}(\tilde{Y}(m,s,T)) \leq \mathrm{MSE}(\bar{Y}(m,s,T))$ when $p < q$, as was suggested above.

We conclude this section by showing $b \geq \sigma^2$. We first observe that $b(x)$ solves Poisson's equation

$$b(x) - E_x b(X(1)) = f_c(x),$$

where $f_c(x) = f(x) - r$. Additionally, $E_\pi b(X(0)) = 0$. Then,

$$\sum_{k=0}^{n-1} f_c(X(k)) = \sum_{k=1}^{n+1} D_k + b(X(0)) - b(X(n+1))$$

where $D_k = b(X(k)) - E\{b(X(k))|X(k-1)\}$ are martingale differences. Note that if $X(0) \stackrel{\mathcal{D}}{=} \pi$, we can apply the martingale central limit theorem (see p. 205 of BILLINGSLEY (1968)) to conclude that

$$n^{-1/2} \sum_{k=0}^{n-1} f_c(X(k)) \Rightarrow \lambda N(0,1) \tag{3.18}$$

where $\lambda^2 = E_\pi D_1^2$. (The function $b(\cdot)$ is bounded under (2.15).) If the left-hand side of (3.18) is appropriately uniformly integrable, then

$$n^{-1} \mathrm{var}_\pi \sum_{k=0}^{n-1} f_c(X(k)) \to \lambda^2 \tag{3.19}$$

as $n \to \infty$. But

$$\mathrm{var}_\pi \sum_{k=0}^{n-1} f_c(X(k)) = n^2 \cdot \mathrm{var}_\pi \frac{1}{n} \sum_{j=0}^{n-1} Y(j).$$

From (2.17) and (3.19), it follows that $\lambda^2 = E_\pi D_1^2 = \sigma^2$. But D_1 is orthogonal to $b(X(0))$, being a martingale difference, and hence

$$E_\pi b(X(1))^2 = E_\pi D_1^2 + E_\pi (E\{b(X(1))|X(0)\}^2).$$

Since $b = E_\pi b(X(0))^2$, it is evident that $b \geq \sigma^2$.

4. Conclusions

The non-rectangular sampling plan introduced in this paper has a lower mean square error than that of the corresponding rectangular plan that involves an equivalent amount of computer time, when the "initial transient" decays slowly. This, of course, is precisely the setting in which the method of multiple replicates exhibits its poorest behavior (relative to a single replicate method). Thus, the non-rectangular plan described here is most beneficial in precisely those problems for which multiple replicates is typically most ineffective.

It should be clear that the replication component of this non-rectangular plan is well-suited to parallel computation. However, the generation of the initial condition $X_0(s)$ is not easily adapted to the parallel setting. This aspect of the sampling plan described here deserves further attention.

Finally, it should be mentioned that a great deal of empirical work remains to be done in understanding the advantages and limitations of this non-rectangular method, when applied to "real world" problems.

References

1. BILLINGSLEY, P. (1968). *Convergence of Probability Measures*, John Wiley, New York.
2. KELTON, W. D. (1986). Replication splitting and variance for simulating discrete-parameter stochastic processes. *Operations Research Letters* 4, 275–279.
3. NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, Cambridge, UK.