VARIANCE REDUCTION IN MONTE CARLO SIMULATION

BY

MARK BROWN, HERBERT SOLOMON and MICHAEL A. STEPHENS

TECHNICAL REPORT NO. 274
JUNE 26, 1979

PREPARED UNDER CONTRACT
NO0014-76-C-0475 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Variance Reduction in Monte Carlo Simulation

By

Mark Brown, Herbert Solomon and Michael A. Stephens

TECHNICAL REPORT NO. 274

JUNE 26, 1979

Prepared under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Partially supported under U.S. Army Research Office Grant DAAG-29-77-G-0031
and issued as Technical Report No. 35.
Variance Reduction in Monte Carlo Simulation

by

Mark Brown
Herbert Solomon
and
Michael A. Stephens

1. Introduction.

Monte Carlo simulation is employed in a large variety of problems. Frequently, one is interested in the expectation of a function \( g(X_1, \ldots, X_N) \) where \( < X_i, i > 1 > \) is i.i.d. with known distribution \( F \) and \( N \) is a stopping time (often a constant). The procedure followed is to generate a large number of samples \( (X_1^{(i)}, \ldots, X_N^{(i)}), i = 1, 2, \ldots, M, \) and estimate the expectation of interest by

\[
\frac{1}{M} \sum_{i=1}^{M} g(X_1^{(i)}, \ldots, X_N^{(i)}).
\]

An interesting aspect of the simulation estimation problem is that \( F \) is known. Thus functions of the form \( \ell(F, X_1, \ldots, X_N) \) can be employed as estimators, while in statistical estimation problem with \( F \) unknown \( \ell \) cannot be computed from the data and is thus not considered to be an estimator. Thus the class of estimators is considerably wider in Monte Carlo problems.

One approach available to reduce the variance of the Monte Carlo estimator is to find a function \( \ell(F, X_1, \ldots, X_N) \) with the same expectation as \( g \), and with smaller variance. Then \( \ell \) rather than \( g \) is averaged over the \( M \) samples. Of course, \( \ell = E_F g \) fits this description but were
it directly computable one would not need to simulate in the first place. Thus an important requirement of \( F \) is that it be simply computable.

We illustrate the above remarks by considering the problem of Monte Carlo estimation of \( M(t) = \text{EN}(t) \), the expected number of renewals in \([0, t]\) for a renewal process with known interarrival time distribution \( F \). Several unbiased estimators which compete favorably with the naive estimator, \( N(t) \), are presented and studied.

We believe that our approach and methodology, although only applied to renewal function estimation in this paper, can be useful in a large variety of Monte Carlo simulation problems.

2. Assume that \( \langle X_i, i \geq 1 \rangle \) is i.i.d. with cdf \( F \) where \( F(0) = 0 \). Define \( S_0 = 0, S_n = \sum_{i=1}^{n} X_i, n = 1, 2, \ldots, N(t) = \max(n: S_n \leq t) \), and \( M(t) = EN(t), t \geq 0 \). Sometimes we consider the point \( t = 0 \) as a renewal epoch. In this case we use \( N_0(t) = N(t) + 1 \) and \( M_0(t) = M(t) + 1 \). The renewal age at time \( t \) is defined by \( A(t) = t - S_{N(t)} \); \( \Pr(A(t) = t) = \overline{F}(t) \) and \( d\overline{F}(x) = \overline{F}(x) dM(t-x) \) for \( 0 \leq x < t \), thus \( d\overline{F}(x) = \overline{F}(x) dM_0(t-x) \) for \( 0 \leq x \leq t \).

Define

\[
\bar{\delta}_i = \begin{cases} 
1 & \text{if } S_i \leq t \\
0 & \text{if } S_i > t 
\end{cases}
\]

Then \( N(t) = \sum_{i=1}^{\infty} \bar{\delta}_i \) and \( M(t) = E \sum_{i=1}^{\infty} \bar{\delta}_i = \sum_{i=1}^{\infty} F(i)(t) \), where \( F(i) \) is the \( i \)th convolution of \( F \).
To estimate $F^{(i)}(t) = E\delta_i$ we will use

$$E(\delta_i | X_1, \ldots, X_i-1) = E(\delta_i | S_i-1) = F(t-S_i-1).$$

We then estimate $M(t)$ by:

$$M_F(t) = \sum_{i=1}^{\infty} F(t-S_i-1) = \sum_{i=1}^{N(t)+1} F(t-S_i-1).$$

Since $\text{Var}(F(t-S_i-1)) = \text{Var}[E(\delta_i | S_i-1)] \leq \text{Var} \delta_i$, we have replaced each component, $\delta_i$, by a component with the same expectation and smaller variance. Intuitively we would expect that if we reduce the variability at each stage (given the past) then we should reduce the variability of the overall estimator. However, the computation of variance involves covariance terms, and if these are increased while variances are decreased there can conceivably be an increase in variance.

Theorem 1 (below) demonstrates that $M_F(t)$ does indeed have lower variance than $M(t)$.

**Theorem 1.** $M_F(t)$ is an unbiased estimator of $M(t)$ and $\text{Var} N(t) - \text{Var} M_F(t) = E[2M(A(t))-F(A(t))] \geq 0$, with strict equality if $F(t) > 0$.

Before proving theorem 1 we comment that the reduction in variance is unsatisfactorily small for large $t$. If $\mu_2 = EX^2 < \infty$ then $E[2M(A(t))-F(A(t))] = O(1)$, thus $\text{Var} N(t)$ and $\text{Var} M_F(t)$ are of the form $\gamma t + O(1)$ with common $\gamma$, and we improve only the asymptotically negligible $O(1)$ term. Estimators considered in later sections do considerably better for large $t$. 

3
Proof of Theorem 1. Express $M_F(t)$ as

$$F(t) + \int_0^t F(t-x) dN(x) = \int_0^t F(t-x) dN_0(x).$$

Then

$$EM_F(t) = \int_0^t F(t-x) dM_0(x) = \int_0^t 1 dM_0(x) - \int_0^t \bar{F}(t-x) dM_0(x)$$

$$= M_0(t) - \int_0^t \frac{dF(t-x)}{A(t)} = M_0(t) - 1 = M(t).$$

Now,

$$EM_F^2(t) = \int_0^t F^2(t-x) dM_0(x)$$

$$+ 2 \iint_{r \leq s} F(t-r)F(t-s) dM_0(r)dM_0(s-r).$$

We evaluate this expression in several steps:

(i) \( \int_0^t F^2(t-x) dM_0(x) = \int_0^t F(t-x) dM_0(x) - \int_0^t F(t-x) \bar{F}(t-x) dM_0(x) \)

\( = M(t) - EF(A(t)). \)

(ii) \( F(t-r)F(t-s) = 1 - \bar{F}(t-r) - \bar{F}(t-s) + \bar{F}(t-r)\bar{F}(t-s). \)

(iii) \( 2 \iint_{r \leq s} 1 dM_0(r)dM_0(s-r) = 2 \int_0^t M(t-r)dM_0(r) = 2M(t) + 2M^{(2)}(t). \)

(iv) \( -2 \iint_{r \leq s} \bar{F}(t-r)dM_0(r)dM_0(s-r) = -2 \int_0^t \bar{F}(t-r) M(t-r)dM_0(r) \)

\( = -2EM(A(t)). \)
(v) \[-2 \iiint_{r<s} \overline{F}(t-s) \, dM_0(r) \, dM_0(s-r) = -2 \int_{r=0}^{t} F(t-r) \, dM_0(r) = -2M(t) .\]

(vi) \[2 \iiint_{r<s} \overline{F}(t-r) \overline{F}(t-s) \, dM_0(r) \, dM_0(s-r) = 2 \int_{r=0}^{t} F(t-r) \overline{F}(t-r) \, dM_0(r) = 2EF(A(t)) .\]

Combining (i)-(vi) we obtain:

(2) \[EM_F^2(t) = M(t) + 2M^{(2)}(t) - E(2M(A(t)) - F(A(t))) .\]

Furthermore

(3) \[EN^2(t) = E\left[ \int_0^t 1dN(t) \right]^2 = M(t) + 2 \iiint_{r<s} dM(r) \, dM(s-r) = M(t) + 2M^{(2)}(t) .\]

Thus from (2) and (3):

\[\text{Var} \, N(t) - \text{Var} \, M_F(t) = E[2M(A(t)) - F(A(t))] .\]

Since

\[M(s) = \sum_{i=1}^{\infty} F^{(i)}(s), \quad 2M(s) - F(s) = F(s) + 2 \sum_{i=2}^{\infty} F^{(i)}(s) \geq 0 ;\]
thus \( E[2M(A(t)) - F(A(t))] \geq 0 \) for all \( t \) and is strictly positive for \( F(t) > 0 \).

3. In this section we assume that \( F \) is continuous. The cumulative hazard \( H \) is defined by \( H(t) = -\log F(t) \). When \( F \) is absolutely continuous with density \( f \) then \( H(t) = \int_0^t h(y)dy \) where \( h \) is the hazard function, \( h(t) = \frac{f(t)}{F(t)} \).

Our next estimator is based on the intuitive idea that \( E(dN(s)|\text{past}) = dH(A(s)) \). Thus instead of using \( N(t) = \int_0^t dN(s) \) we try

\[
M_H(t) = \int_0^t dH(A(s)) = \sum_{1}^{N(t)} H(X_i) + H(A(t)) = \sum_{1}^{N(t)+1} H_i
\]

where \( H_i = H[(t-S_{i-1}) \wedge X_i] \) (where \( a \wedge b = \min(a,b) \)).

Note that \( N(t) = \sum_{1}^{\infty} \delta_i \) while \( M_H(t) = \sum_{1}^{\infty} H_i \). Thus \( \delta_i \) is replaced by \( H_i \), and \( E(\delta_i|S_{i-1}) = E(H_i|S_{i-1}) = F(t-S_{i-1}) \).

The process \( M_H(t) \) is a cumulative process in the sense of Smith [3]. Thus (Smith [3])

\[
\text{Var } M_H(t) \sim \frac{t}{\mu} E[H(X)] - \left( \frac{EH(X)}{\mu} \right)^2,
\]

where \( \mu = EX \). But \( H(X) = -\log F(X) \) is exponentially distributed with parameter 1, thus:

\[
E[H(X) - \frac{EH(X)}{EX} X]^2 = 1 + \frac{\sigma^2}{\mu^2} - \frac{2\sigma}{\mu}.
\]
where $\rho$ is the correlation coefficient between $X$ and $H(X)$ and $\sigma^2$ is the variance of $X$. Thus $M_H(t)$ is asymptotically better than $N(t)$ for $\rho > \mu/2\sigma$, asymptotically worse than $N(t)$ for $\rho < \mu/2\sigma$.

In general if we have two unbiased estimators of a parameter, $T_1$ and $T_2$, with covariance matrix $A$, then the minimum variance unbiased estimator of the form $\alpha T_1 + (1-\alpha)T_2$ is the one with

$$\alpha = \frac{\sum_{j=1}^{2} A_{1j}^{-1}}{\sum_{i=1}^{2} \sum_{j=1}^{2} A_{ij}^{-1}}.$$ 

The variance of this estimator is

$$\frac{1}{\sum_{i,j} A_{ij}^{-1}}.$$ 

The idea now is to let $A$ be the asymptotic covariance matrix of

$$\left(\frac{N(t)}{\sqrt{t}}, \frac{M_H(t)}{\sqrt{t}}\right)$$

and to employ the above result to obtain an unbiased estimator which improves on both $M_H(t)$ and $N(t)$ for large $t$. We already know the $O(t)$ terms for $\text{Var} N(t)$ and $\text{Var} M_H(t)$. We only need the leading term for $\text{Cov}(N(t), M_H(t))$. This is given in lemma 1 below.

Lemma 1. If $\sigma^2$ is finite then

$$\text{Cov}(N(t), M_H(t)) = \frac{t}{\mu} \left( \frac{\sigma^2}{\mu^2} - \frac{\sigma^2}{\mu} \right) + o(t).$$
Proof.

\[
\text{Var}(N(t) - M_H(t)) = \text{Var} \sum_{i=1}^{\infty} (\delta_i - H(t-S_{i-1}) \land X_i)
\]

\[
= \sum_{i=1}^{\infty} \text{E} \text{Var}[\delta_i - H(t-S_{i-1}) \land X_i | S_{i-1}] = \text{E} \sum_{i=1}^{\infty} F(t-S_{i-1}) = \text{EN}(t) = M(t).
\]

Thus

\[
M(t) = \text{Var}(N(t) - M_H(t)) = \text{Var} N(t) + \text{Var}(M_H(t)) - 2\text{Cov}(N(t), M_H(t))
\]

and therefore

\[
\text{Cov}(N(t), M_H(t)) = \frac{1}{2} [\text{Var} N(t) + \text{Var} M_H(t) - M(t)]
\]

\[
= \frac{t}{2\mu} \left[ \frac{\sigma^2}{\mu^2} + \frac{\sigma^2}{\mu^2} + 1 - \frac{2\sigma}{\mu} - 1 + o(1) \right]
\]

\[
= \frac{t}{\mu} \left( \frac{\sigma^2}{\mu^2} - \frac{\sigma}{\mu} \right) + o(t).
\]

Now

\[
A = \frac{1}{\mu} \begin{pmatrix}
\frac{\sigma^2}{\mu^2} & \frac{\sigma^2}{\mu^2} - \frac{\sigma}{\mu} \\
\frac{\sigma^2}{\mu^2} - \frac{\sigma}{\mu} & \frac{\sigma^2}{\mu^2} + 1 - \frac{2\sigma}{\mu}
\end{pmatrix}
\]

\[
\alpha = \frac{\sum_{j=1}^{2} \sum_{i=1}^{2} A^{-1}_{ij}}{\sum_{i=1}^{2} \sum_{j=1}^{2} A^{-1}_{ij}} = 1 - \frac{\sigma}{\mu}
\]

and

8
\[
\frac{1}{\sum_{i,j} A_{ij}^{-1}} = \frac{\sigma^2}{\mu^2} (1-\rho^2).
\]

Note that the asymptotic relative savings in variance is \(\rho^2\) the square of the correlation coefficient between \(X\) and \(H(X)\). Summarizing:

**Theorem 2.** The estimator

\[
M^*(t) = (1 - \frac{\sigma\rho}{\mu}) N(t) + \frac{\sigma\rho}{\mu} M_H(t)
\]

is an unbiased for \(M(t)\) with variance

\[
\frac{t\sigma^2}{\mu^2} (1-\rho^2) + o(t)
\]

(\(\rho\) is the correlation coefficient between \(X\) and \(H(X)\)). It follows that:

\[
\frac{\text{Var} N(t) - \text{Var} M^*(t)}{\text{Var} N(t)} = \rho^2 + o(1).
\]

**Example:** Let \(H(x) = x^2\), \(F(x) = e^{-x^2}\). Then,

\[
\mu = \int_0^\infty e^{-x^2} dx = \sqrt{\frac{\pi}{2}} \int_{-\infty}^\infty \frac{1}{\sqrt{\pi}} e^{-x^2} dx = \frac{\sqrt{\pi}}{2};
\]

\[
EX^2 = 2 \int_0^\infty xe^{-x^2} dx = 1,
\]

thus

\[
\sigma^2 = 1 - \frac{\pi}{4} = \frac{4-\pi}{4} \quad ; \quad \rho = \frac{1}{\sigma} \left[ \int 2x e^{-x^2} dx - \mu \right] = \frac{\mu}{2\sigma} = \frac{1}{2} \sqrt{\frac{\pi}{4-\pi}}, \quad \rho^2 = \frac{\pi}{4(4-\pi)} = .915.
\]
Thus in this case (Weibull with shape parameter 2) the unbiased estimator $M^*(t)$ has an asymptotic relative reduction in risk over $N(t)$ of 91.5 percent. 

Integration by parts shows that

$$\rho = \frac{1}{\sigma} \int_0^\infty H(x) \tilde{F}(x) dx,$$

since $H(x) = -\log \tilde{F}(x)$ the integral can probably be given an entropy interpretation. Also $\rho = \frac{1}{\sigma} \mathbb{E} \tilde{H}(X)$ where $\tilde{H}(x) = \int_0^x H(z) dz$. This is true since

$$\int_0^\infty H(x) \tilde{F}(x) dx = \int_0^\infty H(x) E I_{X \geq x} dx = E \int_0^\infty H(x) I_{X > x} dx = E \int_0^X H(x) dx = \mathbb{E} \tilde{H}(X).$$

Note that both $\rho$ and $\frac{\sigma \rho}{\mu}$ are invariant under a change of time scale, $t \rightarrow ct$, $c > 0$.

In section 3 we estimated $M(t)$ by a weighted average of $N(t) = \sum_1^\infty \delta_i$ and $M_F(t) = \sum_1^N(t) + 1 H((t-S_{i-1}) \wedge X_1)$. Now we apply the same idea but stagewise. At stage $i$, having observed $X_1, \ldots, X_{i-1}$, $N(t)$ adds the component $\delta_i = I_{X_i \leq t - S_{i-1}}$, while $M_F(t)$ adds $H_i = H((t-S_{i-1}) \wedge X_i)$. Each of $\delta_i$, $H_i$ are conditionally (given $S_{i-1}$) unbiased for $F(t-S_{i-1})$ and unconditionally unbiased for $F^{(i)}(t)$. The approach we now follow is to use the weighted average of $\delta_i$ and $H_i$ which has smallest conditional variance given $X_1, \ldots, X_{i-1}$.
Define $F_i = F(t - S_{i-1})$, $C_i = H(t - S_{i-1})$. Then:

$$\text{Var}(S_i | S_{i-1}) = F_i - F_i^2$$

$$\text{Cov}(S_i, H_i | S_{i-1}) = F_i (F_i - C_i)$$

$$\text{Var}(H_i | S_{i-1}) = F_i + F_i (F_i - 2C_i).$$

The minimum conditional variance (given $X_1, \ldots, X_{i-1}$) unbiased linear combination is then:

$$L_i = (1 - \frac{C_i F_i}{F_i}) S_i + \frac{C_i F_i}{F_i} H_i.$$  

The corresponding estimator of $M(t)$ is:

$$M_L(t) = N(t) - \sum L_i \frac{N(t + 1) H(t - S_{i-1}) F(t - S_{i-1})}{F(t - S_{i-1})} (S_i - H_i).$$

We do not know now $M_L(t)$ compares with the other estimators we have looked at. The variance of an estimator of the form $\sum K_i$ is

$$\sum \text{Var} K_i + 2 \sum_{i < j} \text{Cov}(K_i, K_j);$$

$L_i$ was chosen from among a class of estimators $\sum K_i$ to minimize $\sum \text{Var} K_i$. However we know very little about $\text{Cov}(L_i, L_j)$. This latter quantity must be shown to be suitably small in order to demonstrate that $M_L(t)$ has desirable variance properties.
5. We next consider an unbiased estimator with asymptotic variance $O(1)$. Thus it asymptotically enjoys a 100 percent reduction in variance over $N(t)$.

As is well known $N(t)+1$ is a stopping time and thus by Wald's identity:

$$E\sum_{1}^{N(t)+1}X_i = \mu(M(t)+1)$$

Thus

$$\hat{M}(t) = \frac{S_{N(t)+1}}{\mu} - 1$$

is unbiased for $M(t)$. Now $\text{Var}(S_{N(t)+1}) = \text{Var}(t+Z(t)) = \text{Var} Z(t)$, where $Z(t)$ is the forward recurrence time at $t$. If $\mu^3 = \text{EX}^3 < \infty$ then $\text{Var} Z(t)$ converges to

$$\frac{\mu_3}{3\mu} - \frac{\mu_2^2}{4\mu^2} = \frac{4\mu_3 - 3\mu_2^2}{12\mu^2}$$

as $t \rightarrow \infty$. Thus

$$\text{Var} \hat{M}(t) \rightarrow \frac{4\mu_3 - 3\mu_2^2}{12\mu^4}$$

and is thus $O(1)$. 

12
References


# Variance Reduction in Monte Carlo Simulation

**Authors:** Mark Brown, Herbert Solomon, and Michael A. Stephens

**Performing Organization:** Department of Statistics, Stanford University, Stanford, CA 94305

**Controlling Office:** Office of Naval Research, Statistics and Probability Program Code 436, Arlington, Virginia 22217

**Report Date:** June 20, 1979

This report was partially supported under U.S. Army Research Office Grant DAAG29-77-G-0031 and issued as Technical Report No. 35.

**Keywords:** Monte Carlo simulation, estimation, variance reduction, renewal theory, cumulative processes.

**Abstract:** Please see reverse side.
The problem of Monte Carlo estimation of $M(t) = EN(t)$, the expected number of renewals in $[0,t]$ for a renewal process with known interarrival time distribution $F$, is considered. Several unbiased estimators which compete favorably with the naive estimator, $N(t)$, are presented and studied. We believe that our approach and methodology, although only applied to renewal function estimation in this paper, can be useful in a large variety of problems.