ANTITHETIC VARIATES REVISITED

George S. Fishman and BaoSheng D. Huang

Technical Report 80-4
June 1980

UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL

DISTRIBUTION STATEMENT A
Approved for public release; Distribution Unlimited
ANTITHETIC VARIATES REVISITED

George S. Fishman and BaoSheng D. Huang

Technical Report 80-4
June 1980

Curriculum in Operations Research
and Systems Analysis

University of North Carolina at Chapel Hill

This research was supported by the Office of Naval Research under contract N00014-76-C-0302.

Reproduction in whole or in part is permitted for any purpose of the United States government.
Abstract

This paper extends earlier results in the area of variance reduction techniques applied to simulation on a computer. In particular, it views the antithetic sampling technique as a combination of rotation and reflection sampling on a circle. The covariance structures induced by the techniques separately and together are derived and conditions under which they are optimal sampling plans are described. Rates of convergence for the variance of the sample mean are given for bounded, continuous and discrete random variables and for unbounded continuous random variables with special, although commonly encountered, structure.

The advantage of reflection (basic antithetic) sampling is greatest when a certain symmetry property holds. Rotation-reflection sampling is superior to rotation sampling alone for continuous functions. In the bounded continuous case, convergence is faster with rotation-reflection sampling. In the unbounded continuous case, the results show that rotation-reflection sampling speeds convergence to the large sample convergence rate achievable with rotation sampling alone. For the discrete case, rotation sampling does as well with regard to convergence as rotation-reflection sampling does. However, analysis of the discrete case shows that a sample size $n$ may be considerably better than another $n'$ although $n' > n$. 
1. Introduction

Among statistical topics that arise in computer-based simulation experimentation, variance reduction has long occupied a central position, conceptually if not in practice. Variance reduction denotes the objective of adding procedures to an experiment that allow one to obtain a specified accuracy for less cost than one can achieve in their absence. Conversely, for a specified cost, a variance reduction technique enables one to estimate parameters more accurately than one can without such a technique. The subject is not new, topical publications having appeared over twenty years ago (Hammersley and Morton (1956), Hammersley and Mauldon (1956), and Handscomb (1958)). Most textbooks on simulation acknowledge the relevance of the issue (e.g., Emshoff and Sisson (1970), Fishman (1973), Fishman (1978), Gordon (1969) and Naylor et al. (1965)).

Unfortunately, attempts to implement this noble concept in practice have produced few documented cases of success. One notable exception is Carter and Ignall (1975). No doubt, a principal reason for the paucity of success arises from the limited development of the variance reduction technique that has appeared in scholarly journals beyond the original conceptualization of Hammersley and colleagues. However, evidence of change is in the air. Recently Lavenberg, Moeller and Sauer (1979) have attempted to broaden and deepen this development with regard to the variance reduction technique known as the control variate method as it applies to discrete event simulation. They enumerate the do's and don'ts of the method with examples that should prove helpful to potential users. Also, Schruben and Marginin (1978) describe random number stream manipulation techniques designed to induce variance reduction.
The purpose of the present paper is to extend the development of
the antithetic variate method of variance reduction, a procedure first
described in Hammersley and Morton (1956). Our results considerably
augment those of earlier work in this area and were motivated by obser-
vations made in Fishman (1979), which described an application of anti-
thetic variates to population growth simulations. Two examples illustrate
conceptually the value of the method of antithetic variates. Firstly,
consider the evaluation of the integral
\[ \phi = \int_0^1 g(x) \, dx \]
where
\[ \int_0^1 g^2(x) \, dx < \infty. \]
If an analytical solution is unavailable, one can turn either to numerical
integration or to the Monte Carlo method. Let \( U_1, \ldots, U_n \) denote a
sequence of independent observations from the uniform distribution on
\([0,1]\). Let \( U(0,1) \) denote this distribution. Then an unbiased estimator
of \( \phi \) is
\[ \hat{\phi}_n = \frac{1}{n} \sum_{j=1}^{n} g(U_j) \]
with
\[ \text{var} \hat{\phi}_n = O(1/n) \]
so that the standard error of \( \hat{\phi}_n \) decreases as \( O(1/n^{1/2}) \).
This random sampling is the most elementary application of the Monte
Carlo method. Variance reduction techniques denote the use of more advanced sampling designs intended to speed the convergence of $\text{var} \, \hat{\phi}_n$. In particular, the method of antithetic variates aims at inducing a joint distribution among $U_1, \ldots, U_n$ for which

$$\text{var} \, \hat{\phi}_n = o(1/n)$$

while preserving the marginal distributions as $U(0,1)$, which guarantees the unbiasedness of $\hat{\phi}_n$.

As a second example, one may wish to apply variance reduction techniques to a discrete event simulation. Consider the single server queue with i.i.d. interarrival times $A_1, A_2, \ldots$, i.i.d. service times $S_1, S_2, \ldots$, $\{A_i\}$ and $\{S_i\}$ independent, and mean waiting time $\mu$. From Lindley's equation one has for the waiting time of the $i$th completion

$$W_i = \max(0, W_{i-1} + S_i - A_i) \quad i = 1, \ldots, m$$
on a run terminated after $m$ completions. Also, let

$$W_{i,j} = \max(0, W_{i-1,j} + S_{i,j} - A_{i,j}) \quad j = 1, \ldots, n$$
denote the waiting time of completion $i$ on the $j$th of $n$ replications. As an estimator of $\mu$ one has

$$\hat{\mu}_{m,n} = \frac{1}{n} \sum_{j=1}^{n} \bar{W}_j$$

where

$$\bar{W}_j = \frac{1}{m} \sum_{i=1}^{m} W_{i,j}$$.
If replications are independent, then $\text{var} \, \hat{\mu}_{m,n} = O(1/mn)$. One possible application of the method of antithetic variates might be to induce joint distributions among $A_{i,1}, \ldots, A_{i,n}$ and among $S_{i,1}, \ldots, S_{i,n}$ so that

$$\text{var} \, \hat{\mu}_{m,n} = O(1/m) \circ (1/n).$$

Here entries within a column of the array

$$
\begin{array}{cccc}
A_{1,1} & A_{2,1} & \cdots & A_{m,1} \\
A_{1,2} & A_{2,2} & \cdots & A_{m,2} \\
\vdots & \vdots & \ddots & \vdots \\
A_{1,n} & A_{2,n} & \cdots & A_{m,n}
\end{array}
$$

are correlated, but entries within a row are independent. A similar characterization applies to service times.

At this point, it is important to recognize that the direct application of the antithetic method to be described here does not necessarily achieve $O(1/n)$ in $\text{var} \, \hat{\mu}_{m,n}$, a well-known fact in multivariate Monte Carlo sampling. For example, see Hammersley and Handscomb (1964). However, a comprehensive understanding of how the technique works in univariate problems is a prerequisite to devising methods that will achieve the desired effect in multivariate problems such as the aforementioned queueing simulation.

The formal concept of the antithetic variate method first appeared in Hammersley and Morton (1956). Two subsequent papers, Hammersley and Mauldon (1956) and Handscomb (1958), demonstrated a certain optimal
property of the method. Andréasson (1972) and Andréasson and Dahlquist (1972) introduced the formalisms of group representation as a way of analyzing potential antithetic sampling designs, and Roach (1973) attempted to formalize the topic as a transportation assignment problem. The present paper examines and extends the formulations in these early papers into an account that sheds considerable new light on the antithetic method and how it works. Section 2 reviews the formalisms of the antithetic variate method. Section 3 describes a procedures, based on rotation, for collecting \( n > 2 \) antithetic replications that lead to considerably greater accuracy per unit cost than the traditionally recommended antithetic variate method for \( n = 2 \) allows. It also describes several examples that reveal how this rotation sampling performs in selected situations. Section 4 describes a procedure, based on rotation and reflection, that in certain cases improves on the method of Section 3, and illustrates its application to some of the examples in Section 3. Section 5 describes one circumstance in which the results derived here apply to a single server queueing simulation. Both rotation and reflection sampling designs make clear the value of continued study of these procedures.

2. Basic Antithetic Sampling

Consider the random variables \( \eta_1, \ldots, \eta_n \) and suppose one forms the quantity \( h(\eta_1, \ldots, \eta_n) \) and uses it as an estimate of an unknown quantity \( \phi \). If

\[ n^{-1} E h(\eta_1, \ldots, \eta_n) = \phi \]

the estimator is unbiased. Moreover, a low value for \( \text{var} h(\eta_1, \ldots, \eta_n) \)
relative to $\phi^2$ indicates high reliability for $h(n_1,...,n_n)$ as an estimator of $\phi$. An important subclass of interest is the separable function

$$h(n_1,...,n_n) = h_1(n_1) + h_2(n_2) + ... + h_n(n_n).$$

Given $h_1,...,h_n$, the marginal distributions of $n_1,...,n_n$ and the condition $E \sum_{j=1}^{n} h(n_j) = n\phi$, one can concentrate on choosing a joint distribution for $n_1,...,n_n$ to promote reliability without concern for bias. Working in a different, but related, function space facilitates this choice.

Let $n_j$ have the cumulative distribution function (c.d.f.) $F_j$ with inverse distribution function

$$G_j(x) = \inf[y : F_j(y) \geq x, \ 0 \leq x \leq 1].$$

Let $U_1,...,U_n$ denote uniform deviates and define

$$g_j(U_j) = h_j[G_j(U_j) = h_j(n_j).$$

Then the estimator of $\phi$ of interest is

$$T_n = \frac{1}{n} \sum_{j=1}^{n} g_j(U_j). \quad (1)$$

One can now restate the variance reduction problem: Given $g_1,...,g_n$ with $E g_j(U_j) = \phi$, choose the joint distribution of $U_1,...,U_n$ to minimize $\text{var} \ T_n$.

At this point the Antithetic Variate Theorem becomes salient.
Theorem 1. Define $\Omega$ as the set of all functions for which

i. $\omega(z)$ is a 1-1 mapping of $(0,1)$ onto itself.

ii. Except at a finite number of points $z$, $d\omega/dz = 1$.

Also, $I_n \equiv \inf \text{var} T_n$ over all possible stochastic and functional dependences among $U_1, \ldots, U_n$. Then

$$\inf_{\omega_j \in \Omega} \text{var} \left[ \frac{1}{n} \sum_{j=1}^{n} g_j(\omega_j(U)) \right] = I_n.$$  \hfill (2)

For bounded $g_1, \ldots, g_n$, Hammersley and Mauldon (1956) give the proof for $n = 2$ and Handscomb (1956) gives the proof for $n \geq 2$. Recently Wilson (1979) has extended the theorem to unbounded $g_1, \ldots, g_n$.

Theorem 1 has profound implications. It says that one can achieve the infimum $I_n$ by generating a uniform deviate $U$ and applying measure preserving transformations on $(0,1)$. As an example, consider the case of $n = 2$, $h_1(x) = h_2(x)$ and monotone, $g_1(x) = g_2(x)$, and $G_2(y) = G_1(1 - y)$. Then the sampling design $\omega_1(U) = \omega_2(U) = U$ gives

$$T_2 = \frac{1}{2} [h_1(G_1(U)) + h_2(G_2(U))]$$

$$= \frac{1}{2} [g_1(U) + g_1(1 - U)]$$ \hfill (3)

for which $\text{var} T_2 = I_2$. In the case $h_1(x) = x$, $g_1 = G_1$, the minimal variance implies that no other method of generating $\eta_1$ and $\eta_2$ produces a more negative correlation, a result well known in probability theory. See Hoeffding (1940), Fréchet (1951), Mardia (1976) and Whitt (1976).

It is this form of basic antithetic sampling ($\omega_1(U) = \omega_2(U)$) that...
textbooks on simulation usually describe.

The problem that now arises is to choose \{\omega_j(U); j = 1,\ldots,n\} that achieves the infimum of \(\text{var} \, T_n\) for \(n > 2\). This is not a simple problem nor do we claim to have solved it entirely. However, our results are encouraging. Section 3 describes the concept of rotation sampling and shows its optimality under specified conditions. Section 4 then combines basic antithetic sampling with rotation sampling into a rotation and reflection sampling scheme that considerably accelerates the convergence of \(\text{var} \, T_n\).

3. Rotation Sampling

The task of selecting among alternative measure preserving transformations on \([0,1)\) can be simplified at the outset by considering two particular sets. Firstly, consider transformations of the form

\[
\omega(U) = U \quad 0 \leq U < \beta \\
= 1 + \beta - U \quad \beta \leq U < 1 \quad 0 < \beta < 1 .
\]

Figure 1a shows an example. Since these fail to satisfy point ii of Theorem 1, we omit them from further consideration. As a second alternative, consider the transformations

\[
\omega(U) = U \quad 0 \leq U < \beta_1 \\
= U + \beta_3 - \beta_2 \quad \beta_1 \leq U < \beta_2 \\
= U + \beta_1 - \beta_2 \quad \beta_2 \leq U < \beta_3 \\
= U \quad \beta_3 \leq U < 1
\]
for fixed \(0 < \beta_1 \leq \beta_2 \leq \beta_3 < 1\). Note that \(\omega\) covers the unit interval in non-overlapping segments. Figure 1b shows an example. Observe that mappings of this form have several constants to be evaluated, thereby adding to the selection problem.

Here we confine our attention to the set of one-parameter transformations

\[
\omega_j(U) = U \circ \theta_j = U + \theta_j \quad 0 \leq U < 1 - \theta_j \\
= U + \theta_j - 1 \quad 1 - \theta_j \leq U < 1 \quad (4)
\]

for \(j = 1, \ldots, n\). Since these transformations constitute rotations on the unit circle, we refer to (4) as rotation sampling. For convenience of exposition, assume that a) \(\eta_1, \ldots, \eta_n\) have common c.d.f. \(F\) with corresponding inverse distribution function \(G\) and b) \(h_1(x) = \ldots = h_n(x) = h(x)\). Since \(g_j = h_j G_j\), it follows that \(g_1 = \ldots = g_n = g\).
Lastly, assume that $c) \int_0^1 g^2(u) \, du < \infty$.

One can now write (1) as

$$T_n = \frac{1}{n} \sum_{j=1}^{n} g(U \Theta \theta_j)$$

(5)

and let

$$P(\theta) = E g(U) g(U \Theta \theta) - \phi^2$$

(6)

Among the properties that follow from (5) and (6) are

Property 1. (unbiasedness) $E g(U \Theta \theta_j) = E g(U) = \phi$.

Property 2. (continuity) The function $P$ is continuous on $(-\infty, \infty)$.

Property 3. (differentiability). If $g$ is continuous on $[0,1]$, then $P$ is differentiable on $[0,1]$. If $g$ is continuous but unbounded at $u = 1$, then $P$ is differentiable on $(0,1)$. If $\eta_1, \ldots, \eta_n$ have a discrete or mixed marginal distribution, then $P$ has nondifferentiable points in $(0,1)$.

Property 4. (periodicity) $P(\theta) = P(\theta \mod 1) \quad \theta \in (-\infty, \infty)$.

Property 5. (symmetry about $\theta = 1/2$) $P(\theta) = P(1 - \theta)$.

Property 6. (symmetry about $\theta = 0$) $P(-\theta) = P(\theta)$.

Property 7. (exhaustiveness) $\int_0^1 P(\theta) \, d\theta = 0$.

Property 8. (stationarity)

$$\text{cov}[g(U \Theta \theta_j), g(U \Theta \theta_k)] = E g(U \Theta \theta_j) g(U \Theta \theta_k) - \phi^2 = P(\theta_j - \theta_k)$$

$0 \leq \theta_j, \theta_k$.

Property 9. (upper bound) $P(0) = P(1) > P(\theta) \quad \theta \in (0,1)$.

Property 10. (lower bound) If $P$ is convex on $[0,1]$, $P(1/2) \leq P(\theta)$. 
These properties, especially those relating to stationarity and symmetry, prove useful when selecting $\theta_1, \ldots, \theta_n$ to minimize $\text{var} \ T_n$.

This problem can now be formulated as

$$\begin{align*}
\text{minimize} \quad & V(\theta_1, \ldots, \theta_n) = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} P(\theta_j - \theta_k) \\
\text{subject to} \quad & 0 \leq \theta_1, \\
& \theta_j \leq \theta_{j+1} \quad j = 1, \ldots, n-1, \\
& \theta_n \leq 1.
\end{align*}$$

Expression (7) is equivalent to minimization of the average correlation coefficient of $h(n_1), \ldots, h(n_n)$. This formulation leads to Theorem 2.

**Theorem 2.** If $C$ is a convex function on $[0,1]$ and symmetric about $z = 1/2$, then for given $n \geq 2$, $z^* = (1/n, \ldots, 1/n)$ is an optimal solution of the optimization problem:

$$\begin{align*}
\min_{\xi = (z_1, \ldots, z_{n-1})} \quad & w(\xi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} C(z_i + \ldots + z_{j-1}) \\
\text{subject to} \quad & \sum_{i=1}^{n-1} z_i \leq 1, \\
& 0 \leq z_i \quad i = 1, \ldots, n-1.
\end{align*}$$

See the Appendix for the proof.

Letting $C = P$ tells us that the assignment $\theta_j^* = \sum_{k=1}^{j-1} z_k^* = (j-1)/n$ for $j = 1, \ldots, n$ gives

$$V(\theta_1^*, \ldots, \theta_n^*) \leq V(\theta_1, \ldots, \theta_n).$$

The assignment $\theta_n^* = (\theta_1^*, \ldots, \theta_n^*)$ leads to considerable convenience.
In particular, \( \{ \omega_j^*(U) = U \otimes \theta_j^*; j = 1, \ldots, n \} \) form a finite cyclic abelian group. Define \( P_{i-j} = P(\theta_i^* \otimes \theta_j^*) \). Then \( \{ g(U \otimes \theta_j^*) \} \) has the covariance matrix

\[
P_n = \begin{pmatrix}
P_0 & P_1 & P_2 & \cdots & P_2 & P_1 \\
P_1 & P_0 & P_1 & \cdots & P_3 & P_2 \\
P_2 & P_1 & P_0 & P_1 & \cdots & P_4 & P_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
P_2 & P_3 & P_4 & \cdots & P_0 & P_1 \\
P_1 & P_2 & P_3 & \cdots & P_0 & P_1 \\
\end{pmatrix}
\]

Here row \( j+1 \) is row \( j \) with elements shifted one position to the right and the right-most entry in row \( j \) assigned to the left-most position in row \( j+1 \). A matrix with this property is called a circulant. Its \( k \)th eigenvector is \( \{ e^{i2\pi jk/n}; j = 0, \ldots, n-1 \} \) where \( i = \sqrt{-1} \), which gives the eigenvalues

\[
\tau_{k,n} = \sum_{j=0}^{n-1} P_j e^{2\pi i(k-1)j/n} \quad k = 1, \ldots, n .
\]

In particular, note that the unitary matrix \( V_n = \| e^{2\pi i(k-1)j/n} \| \) orthogonalizes \( P_n \) and that \( \tau_{k,n} = \tau_{n-k+2,n} \) for \( k = 2, \ldots, n \).

Regardless of whether or not \( \vartheta_n^* \) is optimal, the resulting symmetry in \( P_n^* \) affords an understanding of the rate of convergence of \( \text{var} T_n^* \) with \( n \) where

\[
T_n^* = \frac{1}{n} \sum_{j=1}^{n} G(U \otimes (j-1)/n) .
\]
Theorem 3. If one uses the transformations \( \{ \omega_j^*; j = 1, \ldots, n \} \), then
\[
\text{var } T_n^* = \frac{1}{n} \sum_{j=0}^{n-1} P(j/n) = \tau_{1,n}/n . \tag{11}
\]

Proof. Observe that
\[
\text{var } T_n^* = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{j=1}^{n} P(k-j)
\]
where the summation is over all elements in \( P_n \). Since each row of \( P_n \) contains the same elements, it follows that the summation over all elements in \( P_n \) is equivalent to
\[
\text{var } T_n^* = \frac{1}{n^2} \left( n \sum_{j=0}^{n-1} P(j) \right) = \frac{1}{n} \sum_{j=0}^{n-1} P(j/n) . \tag{12}
\]
From (10) with \( k = 1 \), it is clear that
\[
\text{var } T_n^* = \tau_{1,n}/n .
\]

The convergence problem now becomes one of showing how \( \tau_{1,n} \) behaves as \( n \to \infty \) under alternative restrictions on \( g \). To put this problem in perspective, observe that
\[
\text{var } T_n^* = \frac{1}{n} \left[ \sum_{j=1}^{n-1} P(j/n) + \frac{1}{2} P(0) + \frac{1}{2} P(1) \right] - \int_0^1 P(\theta) \, d\theta
\]
so that one can interpret \( \text{var } T_n^* \) as the error incurred in using the trapezoidal rule to approximate the integration of \( P \) over \([0,1] \).

Theorem 4. If one uses the transformation \( \{ \omega_j^*; j = 1, \ldots, n \} \) then \( T_n^* \) is the minimum variance unbiased estimator of \( \phi \) for fixed \( n \).
Proof. Consider the estimator

\[ \hat{T}_n = \sum_{j=1}^{n} c_j g(U \theta \theta^*) \]

\[ \sum_{j=1}^{n} c_j = 1 . \]

Let \( c_n = (c_1, \ldots, c_n) \). In order for \( \hat{T}_n \) to be the minimum variance unbiased estimator of \( \phi \), one needs

\[ c_n = (I_n^T P_n^{-1} I_n)^{-1} I_n^T P_n^{-1} \]

where \( I_n \) is an \( n \times 1 \) vector of ones. Since \( V_n \) is a unitary matrix, we have \( V_n^{-1} = V_n^T \) and \( I_n = V_n^T P_n V_n \) where \( I_n \) is a diagonal matrix with \( I_{k,n} \) in row \( k \) and column \( k \). Then

\[ c_n = (I_n^T V_n I_n^{-1} V_n^T I_n)^{-1} I_n^T V_n I_n^{-1} V_n = \frac{1}{n} (1, \ldots, 1) \]

so that \( \hat{T}_n = T_n^* \), which proves the theorem. This result is also noted in Andréasson (1972).

An equivalent representation for \( T_n^* \) proves useful.

**Lemma 5.1.** For the rotation scheme \( \theta_n^* \)

\[ T_n^* = \frac{1}{n} \sum_{j=0}^{n-1} g(\xi_j + \frac{1}{n}) \] (13)

where \( \xi \) is from \( U(0,1) \).
Proof. One has

$$T_n^* = \frac{1}{n} \sum_{j=0}^{n-1} g(U \oplus \frac{j}{n})$$

$$= \frac{1}{n} \left[ \sum_{j=0}^{m} g(U + \frac{j}{n}) + \sum_{j=m+1}^{n-1} g(U + \frac{j-n}{n}) \right]$$

where \( m = \lfloor n(1 - U) \rfloor \). Let \( \xi = 1 - n(1 - U) + m = nU \mod 1 \) so that

$$T_n^* = \frac{1}{n} \left[ \sum_{j=0}^{m} g(\xi + j + n-m-1) \right] + \sum_{j=m+1}^{n-1} g(\xi + j - n)$$

$$= \frac{1}{n} \left[ \sum_{k=0}^{n-m-1} g(\xi + k) + \sum_{k=m+1}^{n-1} g(\xi - k) \right]$$

Clearly \( \xi \) is from \( U(0,1) \). Expression (13) is identical with the Hammersley and Morton (1956) formulation for \( n > 2 \). Their convergence results make use of the Euler summation formula (see Fort 1948, p. 53)

$$\frac{1}{n} \sum_{j=0}^{n-1} q(\frac{x + j}{n}) = \int_0^1 q(t) \, dt + \sum_{k=1}^{m} \frac{B_k(x)[q^{(k-1)}(1) - q^{(k-1)}(0)]}{k! n^k} + o(1/n^m) \quad (16)$$

where \( 0 \leq x < 1 \) and \( B_k(x) \) denotes the kth Bernoulli polynomial for an arbitrary function \( q \) whose first \( m \) derivatives exist. Note that (12) and (13) both are amenable to this representation subject to the existence of the appropriate derivatives.

Bounded \( g \). We now explore the convergence of \( \text{var} T_n^* \) under alternative restrictions on \( g \). Theorem 5 relates to bounded continuous \( g \)
with finite first derivative and Theorem 6, to piecewise linear $g$ with finite discontinuities.

Theorem 5. (Hammersley and Morton 1956). If $g \in C^1[0,1]$, then

\[ a. \quad T_n^* = \frac{\phi - \frac{1}{n} (V - 1/2)(g(1) - g(0))}{\frac{1}{12n^2} [g(1) - g(0)]^2 + o(1/n^2)} + o(1/n) \]

\[ b. \quad \text{var} \ T_n^* = \frac{1}{12n^2} [g(1) - g(0)]^2 + o(1/n^2) \]

where $V = nU \mod 1$ is from $U(0,1)$.

Proof. Result $a$ follows from substitution into (16). Since $T_n^*$ is unbiased, result $b$ follows directly.

To appreciate the significance of this result, one needs a measure of variance reduction. One suggested measure is

\[ VR(\theta_n^*) = \frac{\text{variance without variance reduction technique}}{\text{variance with variance reduction technique}} \]

Then for rotation sampling with bounded $g$

\[ \lim_{n \to \infty} n^{-1} VR(\theta_n^*) = 0(1) \]

so that variance reduction is $O(n)$.

Example 5.1. Consider a Beta random variable with c.d.f. $F(x) = x^\alpha \quad 0 \leq x \leq 1$ and $0 < \alpha \leq 1$ so that $G(U) = U^{1/\alpha}$. Let $g = G$. Then $T_n^* = \frac{\alpha}{\alpha + 1} - (V - 1/2)/n + o(1/n)$ and $\text{var} \ T_n^* = 1/12n^2 + o(1/n^2)$.

Observe that for $0 < \alpha < 1$ the corresponding p.d.f. is unbounded at $x = 0$. For the bounded case ($\alpha > 1$), Theorem 5 does not apply, and one needs an additional result.

Corollary 5.1. If $g$ is continuous on $[0,1]$, then $\text{var} \ T_n^* = o(1/n)$.

Proof. If $g$ is continuous, $P \in C^1[0,1]$. Using (16) with $P$ and $x = 0$, one has

\[ \text{var} \ T_n^* = \int_0^1 P(\theta) \, d\theta + \frac{B_1(0)[P(1) - P(0)]}{n} + o(1/n) \]
Since $P(0) = P(1)$, $\text{var } T^n = o(1/n)$.

A somewhat stronger convergence result than Corollary 5.1 is also possible.

**Corollary 5.2.** If $g$ is continuous on $[0,1]$ and $\int_0^1 g'(x) \, dx < \infty$, then $\text{var } T^n = O(1/n^2)$.


**Example 5.2.** For the Beta case with c.d.f. $F(x) = x^\alpha$ $0 \leq x \leq 1$ and $\alpha > 1$, one has $g'(u) = \frac{1}{\alpha} u^{1/\alpha - 1}$ which is unbounded at $u = 0$. However, $\int_0^1 g'(u) \, du = 1 < \infty$ so that $\text{var } T^n = O(1/n^2)$.

**Theorem 6.** If $g$ is piecewise linear with finite discontinuities, then $\text{var } T^n = O(1/n^2)$. See the Appendix for the proof.

**Example 6.1.** Consider a Bernoulli random variable with inverse distribution function

$$G(U) = \begin{cases} 0 & 0 \leq U \leq 1-p \\ 1 & 1-p < U < 1 \end{cases}$$

and let $g = G$. Then $P(\theta) = (p-\theta)^+ + (p+\theta-1)^+ - p^2$ $0 \leq \theta \leq 1$, where $x^+ = \max(0,x)$. Figure 2 shows $P(\theta)$. Note that $P$ is convex but not differentiable at $\theta = p, 1-p$. Also

$$\text{var } T_n^* = \frac{(np \mod 1)(1 - (np \mod 1))}{n^2} = 0 \quad \text{np = integer}$$

$$\leq \frac{1}{4n^2} \quad \text{always},$$

so that variance reduction is infinite when $np = \text{integer}$ and otherwise
is $O(n)$. Moreover, perusal of Table 1 in Fishman (1979) leads to the conjecture: Given $n$ as the maximal permissible sample size, then using only $n^* = \min\{j: j \text{ mod } 1 \text{ is a minimum, } j = 1, \ldots, n\}$ leads to $\text{var } T_{n^*} \leq \text{var } T_j$ for $j = 1, \ldots, n$. For the more general discrete case, Huang (1980) shows that if $F$ assumes only rational values, there exist $n$'s for which $\text{var } T_n^* = 0$.

Unbounded $g$. Here we use results from generalized function theory. Consider the function

$$q(u) = u^a(1 - u)^b r(u) \quad a, b \leq 0 \quad 0 \leq u \leq 1$$

where $q$ is integrable and the first $m$ derivatives of $r$ exist. Then Lyness and Ninham (1967) give the extended Euler-Maclaurin summation formula.
\[
\frac{1}{n} \sum_{j=0}^{n-1} q\left(\frac{x+j}{n}\right) = \int_0^1 q(u) \, du \\
+ \sum_{j=0}^{m-1} \psi_0^{(j)}(0) \cdot \psi_1(-a-j,x) + (-1)^j \psi_1^{(j)}(1) \cdot \psi_1(-b-j,1-x) \\
\quad + O\left(\frac{1}{n^m}\right) \quad 0 \leq x \leq 1
\] (17)

where \(\psi_0(u) = (1 - u)^b r(u)\), \(\psi_1(u) = u^a r(u)\) and \(\zeta(\cdot, \cdot)\) denotes the generalized Riemann zeta function. We then have:

**Theorem 7.** If \(g = q\) and \(r \in C^1[0,1]\), then \(\text{var} \, T_n^* = O\left(\frac{1}{n^{2(1+a)}}\right)\) if \(a \leq b\) and \(\text{var} \, T_n^* = O\left(\frac{1}{n^{2(1+b)}}\right)\) if \(a > b\).

**Proof.** The result follows directly from (17) with \(m = 1\), \(x = nU \mod 1\) and the fact that \(T_n^*\) is unbiased.

Observe that variance reduction if \(O(n^{1+2a})\) for \(a \leq b\) and \(O(n^{1+2b})\) for \(a > b\), so that the efficiency of rotation sampling increases only if \(a > -1/2\) and \(b > -1/2\). But this is precisely the condition that assures a finite variance for \(g\).

**Example 7.1.** Consider the Pareto distribution with c.d.f \(F(x) = 1 - x^c\) \(c < 0\) and inverse distribution function \(G(U) = (1 - U)^{1/c}\). Also let \(g = G\). This representation corresponds to \(a = 0\) and \(b = 1/c\) in (17) so that \(\text{var} \, T_n^* = O\left(\frac{1}{n^{2(1 + 1/c)}}\right)\) which requires \(c > -2\) \((a > -1/2)\) to achieve a variance reduction.

Other types of unbounded variation are also possible. Consider the representation

\[q(u) = u^a (1 - u)^b r(u) \ln u \quad 0 \leq u \leq 1.\]
If \( q \) is integrable and the first \( m \) derivatives of \( r \) exist, Lyness and Ninham (1967) give the extended Euler-Maclaurin summation formula

\[
\frac{1}{n} \sum_{j=0}^{n-1} q\left(\frac{j}{n}\right) = \int_0^1 q(u) \, du + \sum_{j=0}^{m-1} \left[ e_j(x) \frac{\zeta(-a,x) \ln n}{n^{a+j+1}} + \frac{\zeta(-b,x)}{n^{b+j+1}} \right] + O\left(\frac{1}{n^m}\right)
\]

where \( a, b < 0 \) and the coefficients \( e_j(x) \) are independent of \( n \).

This gives rise to Theorem 8.

**Theorem 8.** If \( q = g \) and \( r \in C^1[0,1] \), then \( \text{var } T^*_n = O\left(\frac{\ln n}{n} + a\right)^2 \) if \( b \geq a \) and \( \text{var } T^*_n = O\left(\frac{1}{n^{2(1+b)}}\right) \) if \( b < a \).

**Proof.** The result follows from substitution into (18) with \( m = 1 \) and \( x = nU \mod 1 \) and from the unbiasedness of \( T^*_n \).

The term \( O\left(\frac{\ln n}{n} + a\right)^2 \) calls for additional study. Observe that

\[
\lim_{n \to \infty} \left(\frac{\ln kn}{(kn)^{1+a}}\right)^2 \left(\frac{1}{\ln n}\right)^2 = \frac{1}{k^{2(1+a)}}.
\]

This implies that for sufficiently large \( n \) \( \text{var } T^*_n = O\left(\frac{1}{n^{2(1+a)}}\right) \) for \( a \leq b \). Here the logarithmic singularity slows the convergence rate for moderate \( n \) but ultimately has no limiting effect.

**Example 8.1.** Let \( g(U) = G(1 - U) = -\ln U \) so that \( g(U) \) is an exponential random variable with unit mean. Here \( a = b = 0 \), \( r(u) = 1 \) and \( \text{var } T^*_n = O\left(\frac{1}{n^{2(1+a)}}\right) \). Again for sufficiently large \( n \) \( \text{var } T^*_n = O\left(\frac{1}{n^2}\right) \), the rate achievable for bounded functions.
4. Rotation and Reflection Sampling

As mentioned in Section 2, the use of $U$ and $1-U$ leads to the most negative correlation between two random variables when $h$ is monotone. However, the exclusive use of rotation overlooks the benefit of this transformation. To investigate this alternative more thoroughly, we consider $n = 2m$ replications, where (4) defines $\omega_1(U), ..., \omega_m(U)$

and

$$G_{j+m}(x) = G_j(1-x) = G(x) \quad j = 1, ..., m \hspace{1cm} (19)$$

$$\omega_{j+m}(U) = \omega_j(U) \quad j = 1, ..., m. \hspace{1cm} (20)$$

Although (4) and (20) induce identical c.d.f.'s on $n_1, ..., n_n$, $n_j$ and $n_{k+m}$ have different joint distributions than the corresponding ones for $n_j$ and $n_k$ for $j, k = 1, ..., m$. Also, (20) meets the requirements of point i of Theorem 1.

Note that $g_j(x) = g_{j+m}(1-x) = g(x)$ for $j = 1, ..., m$. Also note that

$$1 - \omega_j(U) = 1 - (U \oplus \theta_j) = (1 - U) \oplus (1 - \theta_j)$$

$$= 1 - \theta_j - U \quad 0 \leq U \leq 1 - \theta_j$$

$$= 2 - \theta_j - U \quad 1 - \theta_j < U < 1.$$

One can now write $T_n$ as

$$T_n = \frac{1}{m} \sum_{j=1}^{m} f(U \oplus \theta_j) \hspace{1cm} (21)$$

where

$$f(U \oplus \theta_j) = \frac{1}{2}[g(U \oplus \theta_j) + g(1 - (U \oplus \theta_j))]. \hspace{1cm} (22)$$
The first term in the summand of (22) provides for rotation on the unit circle; the second provides for reflection on the unit circle.

Properties of interest include:

Property 11. (symmetry) \( f(x) = f(1 - x) \).

Property 12. (unbiasedness) \( E g(l - \theta_0 \theta - U) = E g(U) = \psi \).

Property 13. (stationarity) \( \text{cov}[g(l - \theta_j \theta - U)g(l - \theta_k \theta - U)] = P(\theta_k - \theta_j) \).

Property 14. (stationarity)

\[
\text{cov}[g(u \theta_j)g(l - \theta_k \theta - U)] = Q(\theta_k - \theta_j) \quad \theta_k \leq \theta_j,
\]

where

\[
Q(\theta_k - \theta_j) = \int_0^1 g(u \theta_j)g(1 - \theta_k \theta - u) \, du - \psi^2
\]

\[
= \int_0^1 g(u \theta)g(1 - u) \, du - \psi^2 \quad \theta = \theta_k - \theta_j \geq 0.
\]

Property 15. (exhaustiveness) \( \int_0^1 Q(\theta) \, d\theta = \int_0^1 Q(1 - \theta) \, d\theta = 0 \).

Property 16. (symmetry) If \( g(u) + g(1 - u) = 2g(1/2) \) for \( 0 \leq u \leq 1 \), \( Q(\theta) = Q(1 - \theta) \) for \( \theta \in [0,1] \).

Property 17. (lower bound) \( Q(0) = Q(1) \geq P(1) \).

Property 18. (continuity) \( Q \) is continuous on \([0,1]\).

To investigate the simultaneous benefit of rotation and reflection, it is convenient to study

\[
R(\theta) = \frac{1}{4} [2P(\theta) + Q(\theta) + Q(1 - \theta)] = \int_0^1 f(u)f(u \theta \theta) \, du - \psi^2
\]

\( \theta \in [0,1] \).
Note that

Property 19. (symmetry) \( R(\theta) = R(1 - \theta) \).

Property 20. (upper bound) \( R(0) = R(1) - R(\alpha) \) for \( \alpha \in [0,1] \).

Property 21. (differentiability) If \( g \in C^1[0,1] \), \( R'(0) = R'(1) = 0 \).

Proof:
\[
R(\theta) = \int_0^{1-\theta} f(u) f(u+\theta)du + \int_{1-\theta}^1 f(u)f(u+\theta-1)du - \frac{\theta^2}{2}
\]
\[
R'(\theta) = f(1-\theta)[f(0) - f(1)] + \int_0^{1-\theta} f(u)f'(u+\theta)du - \int_{1-\theta}^1 f(u)f'(u+\theta-1)du
\]
\[
R'(0) = f(1)[f(0) - f(1)] + \int_0^1 f(u)f'(u)du = 0.
\]

By property 11, \( f(0) = f(1) \) and
\[
\int_0^{1/2} f(u)f'(u)du = \int_{1/2}^1 f(u)f'(u)du
\]
so that \( R'(0) = 0 \). By property 19, \( R'(0) = -R'(1) \). Note that this result does not necessarily apply if \( R'(\theta) \) does not exist everywhere on \([0,1]\).

Property 21 prevents us from deriving a result for \( R \) comparable to Theorem 1. Since \( R'(0) = R'(1/2) = R'(1) = 0 \), \( R \) is not convex. Nevertheless, the choice of \( \theta^*_n \) has highly beneficial properties which the next several theorems describe.
Let
\[
T**_n = \frac{1}{m} \sum_{j=0}^{m-1} f(U \theta j/m)
\]
so that
\[
\text{var } T**_n = \frac{1}{m} \sum_{j=0}^{m-1} R(j/m),
\]
since it is easily seen that \( \{f(U \theta j/m); j = 0, \ldots, m-1\} \) is a cyclic stochastic process.

**Theorem 9.** If property 16 holds, then \( \text{var } T**_n = 0 \).

**Proof.** Since
\[
g(U \theta j/n) + g(1 - (U \theta j/n)) = 2g(1/2) \quad j = 0, \ldots, m-1.
\]
\( T**_n = g(1/2) \) and \( \text{var } T**_n = 0 \).

**Bounded g.**

**Theorem 10.** If \( g \in C^1[0,1] \) then

a. \( T**_n = \phi + o(1/n) \).

b. \( \text{var } T**_n = o(1/n^2) \).

**Proof.** Let \( f(u) = q(u) \) as in (16) and take \( x = V = nU \mod 1 \).

By property 11 \( f(0) = f(1) \) so that results a and b follow immediately.

Note that convergence for rotation-reflection is \( o(1/n^2) \) as compared to \( O(1/n^2) \) for rotation sampling alone.

**Corollary 10.1.** If \( g \in C^2[0,1] \), then \( \text{var } T**_n = O(1/n^4) \).

**Proof.** Expression (16) gives
\[
T**_n = \phi + 4 \frac{B_2(V)[f'(1) - f'(0)]}{n^2} + o(1/n^2)
\]
from which $\text{var} T_n^{**} = O(1/n^4)$. Note that the additional smoothness in $g$ considerably increases the convergence rate.

**Theorem 11.** If $g$ is piecewise linear with finite discontinuities, then $\text{var} T_n^{**} = O(1/n^2)$.

**Proof.** See the proof of Theorem 6 in the Appendix with $f = R$.

**Example 11.1.** Consider the Bernoulli case of Example 6.1. Here $Q(\theta) = (p - \lfloor(1 - \theta) - p\rfloor)^+ - p^2$, from which it follows that $2R(\theta) = (p - \theta)^+ + (p + \theta - 1)^+ - p^2 = P(\theta)$. Therefore, $\text{var} T_n^{**} = (np \mod 1)(1 - (np \mod 1))/n^2 = \text{var} T_n^*$. Also, $R$ is convex so that $\mathbb{Q}_n$ retains its optimal property.

**Unbounded g.**

**Theorem 12.** For $g(u) = u^a(1 - u)^b r(u)$ where $a, b < 0$, $g$ is integrable and $r \in C^1[0,1]$, $\text{var} T_n^{**} = O(1/n^{2(1+a)})$ for $a \leq b$ and $\text{var} T_n^{**} = O(1/n^{2(1+b)})$ for $a > b$.

**Proof.** By appropriate use of Lemma 5.1 one can show that

$$T_n^{**} = \frac{1}{2^m} \sum_{j=0}^{n-1} \left[ g\left(\frac{j}{m}\right) + g\left(\frac{1-j}{m}\right) \right]$$

where $\xi = mU \mod 1$. Using (17) leads to the result. Note the absence of any advantage in terms of the ultimate convergence rate.

**Theorem 13.** For $g(u) = u^a(1-u)^b q(u) \ln u$ where $a, b \leq 0$, $g$ is integrable and $q \in C^1[0,1]$,

a. $\text{var} T_n^{**} = O((\ln n/n^{1+a})^2)$ for $a \leq b < 0$

b. $\text{var} T_n^{**} = O(1/n^{2(1+b)})$ for $b < a$

c. $\text{var} T_n^{**} = O(1/n^2)$ for $a = b = 0$. 
Proof. Results a and b follow directly from (18) and (19) as in Theorem 8. Result b arises as a consequence of \( \zeta(0,x) = -\zeta(0,1-x) \). The implication of result c for rotation-reflection sampling is that the large sample convergence rate \( O(1/n^2) \) is achieved faster than in the case of rotation sampling alone. A reexamination of the exponential case in Example 8.1 illustrates this case.

5. What About Discrete Event Simulation?

In discrete event simulation, the sampling problem incurred usually is a multivariate one for which it is known that the variance reduction properties for the univariate case do not necessarily hold. Although this topic remains for future research, at least one important situation that arises in congestion property establishes the importance of studying the univariate case. Let us return to the single server queue simulation of Section 1. Let \( B_{ij} = S_{ij} - A_{ij} \) for \( i = 1,\ldots,m \) waiting times on replication \( j = 1,\ldots,n \). Then as the traffic intensity approaches unity,

\[
W_{ij} \approx W_{0j} + \sum_{k=1}^{i} B_{kj}
\]

so that

\[
\bar{W}_{ij} \approx W_{0j} + \frac{1}{m} \sum_{i=1}^{m} (m - i + 1)B_{ij}
\]

Here \( \bar{W}_{ij} \) becomes a sum of independent random variables and if one uses rotation-reflection sampling to generate \( B_{i1},\ldots,B_{in} \) for each \( i \) one can expect \( \text{var} \bar{W}_{in} \) to show a convergence rate associated with the known distribution of the \( B_{ij} \). Preliminary sampling experiments confirm this result for large traffic intensities.
6. **Conclusions**

The results presented here extend those in Hammersley and Morton (1956) by showing the covariance structures induced by the rotation and reflection (antithetic) sampling plans, deriving conditions under which these sampling plans are optimal and by examining the unbounded case. For the piecewise linear case, the results suggest that a sample size $n$ can be considerably more desirable than another $n'$ although $n' > n$. The results also show that the benefits of reflection sampling arise principally for symmetric (property 16) functions. The benefit for nonsymmetric unbounded functions is to speed the rate of $\text{var } T_{n}^{**}$ for moderate $n$ to the ultimate rate achievable with rotation sampling alone. This is clearly advantageous when working within a limited budget.
7. References


APPENDIX

Proof of Theorem 2. Lemma A.1 shows that (8) is a convex programming problem. Then we show $z^*$ is a local minimum point. Since a local minimum point of a convex programming problem is also a global minimum point, $z^*$ is a global minimum point.

Lemma A.1. Formulation (8) is a convex programming problem.

Proof. For every $(i,j)$ where $1 \leq i \leq n-1$ and $i+1 \leq j \leq n$, define

$$I_{(i,j-1)}(z) = z_i + \ldots + z_{j-1}$$

where

$$z \in Z = \{ (z_1, \ldots, z_{n-1}) | \sum_{k=1}^{n-1} z_k \leq 1, \; z_k \geq 0 \; \; k = 1, \ldots, n-1 \}.$$ 

Here $Z$ denotes the feasible region of (8), $I_{(i,j-1)}$ is a linear function on $Z$ and

$$C(z_i + \ldots + z_{j-1}) = C(I_{(i,j-1)}(z))$$

is a convex function on $Z$. Since the objective function $w$ in (8) is the sum of convex functions, $w$ is convex on $Z$. Since the constraints in (8) are linear, (8) is by definition a convex programming problem.

Let $m$ and $m'$ be positive integers. Then, convexity gives

$$mc(a) + m'c(b) \geq (m + m')C(\frac{ma + m'b}{m + m'})$$

(A.1)

for any $0 \leq a, b \leq 1$. 
Proof of the Theorem. Since $z^*$ is an interior point of $Z$, there exists an open neighborhood $N(z^*)$ such that

$$z^* \in N(z^*) \subset Z.$$ 

Then $z^*$ is a local minimum point if

$$w(z^*) \leq w(z^* + y)$$

for all $z^* + y$ in $N(z^*)$. The value of the objective function at the perturbed point $z^* + y$ is from (8)

$$w(z^* + y) = w[(z_1^*, ..., z_{n-1}^*) + (y_1, ..., y_{n-1})]$$

$$= w\left(\frac{1}{n} + y_1, ..., \frac{1}{n} + y_{n-1}\right)$$

$$= \sum_{i=1}^{n-1} n-j \sum_{i=1}^{j} c(i_j + \sum_{t=i}^{i+j-1} y_t).$$

Rearranging the terms in (A.3), we have

$$w(z^* + y) = \left[\sum_{i=1}^{n-1} \frac{c(1_n + y_i)}{n} + \sum_{i=1}^{n-1} \frac{c(2_n + y_i)}{n}\right] + \left[\sum_{i=2}^{n-2} \frac{c(2_n + y_i + y_{i+1})}{n} + \sum_{i=1}^{2} \frac{c(n-2_n + \sum_{t=i}^{i+n-1} y_t)}{n}\right] + ...$$

and

$$w(z^*) = \sum_{i=1}^{n-1} \frac{c(1_n)}{n} + \sum_{i=2}^{n-2} \frac{c(2_n)}{n} + \sum_{i=1}^{2} \frac{c(n-2_n)}{n} + ...$$

$$= \left\{ \begin{array}{ll} nc(\frac{1}{n}) + nc(\frac{2}{n}) + ... + nc(\frac{n-1}{2n}) & n \text{ odd} \\ nc(\frac{1}{n}) + nc(\frac{2}{n}) + ... + \frac{n}{2} c(\frac{1}{2}) & n \text{ even} \end{array} \right.$$
The last equality of (A.5) follows from the symmetry of $C$. That is,

$$C(\frac{k}{n}) = C(\frac{n-k}{n}) \quad k = 1, \ldots, n-1.$$  

Now, we show $w(z^* + y) \geq w(z^*)$ by proving that

$$\sum_{i=1}^{n-k} C(\frac{k}{n} + \sum_{t=i}^{i+k-1} y_t) + \sum_{i=1}^{k} C(\frac{n-k}{n} + \sum_{t=i}^{n+i-k-1} y_t) \geq \sum_{i=1}^{n-k} C(\frac{k}{n}) + \sum_{i=1}^{k} C(\frac{n-k}{n}) = nC(\frac{k}{n}) \quad k = 1, \ldots, n-1.$$  

(A.6)

Let $r_{i,k} = \sum_{t=i}^{i+k-1} y_t$. Repeatedly using (A.2), we have

$$\sum_{i=1}^{n-k} C(\frac{k}{n} + r_{i,k}) \geq (n-k)C(\frac{k}{n} + \frac{\sum_{i=1}^{n-k} r_{i,k}}{n-k}).$$  

(A.7)

The symmetry of $C$ gives

$$\sum_{i=1}^{k} C(\frac{n-k}{n} + r_{i,k}) = \sum_{i=1}^{k} C(\frac{k}{n} - r_{i,k}) \geq kC(\frac{k}{n} - \frac{\sum_{i=1}^{k} r_{i,k}}{k}).$$  

(A.8)

The last inequality of (A.8) also follows from repeated use of (A.1).

Combining the results of (A.7) and (A.8), we have

$$\sum_{i=1}^{n-k} C(\frac{k}{n} + r_{i,k}) + \sum_{i=1}^{k} C(\frac{n-k}{n} + r_{i,k}) \geq (n-k)C(\frac{k}{n} + \frac{\sum_{i=1}^{n-k} r_{i,k}}{n-k}) + kC(\frac{k}{n} - \frac{\sum_{i=1}^{k} r_{i,k}}{k}) \geq nC(\frac{k}{n}),$$

which proves the inequality (A.6). From (A.6), (A.5), and (A.4) we have
for all $z^* + y \in N(z^*)$, which completes the proof.

Proof of Theorem 6. The proof follows as a consequence of Lemma A.2.

Lemma A.2. Let $f$ be a continuous piecewise linear function on $[0,1]$ with parameters $s_1, s_2, d_1, d_2$ and $c$ such that

$$f(x) = [s_1 x + d_1]I_{[0,c]}(x) + (s_2 x + d_2)I_{[c,1]}(x)$$

for $0 < c < 1$.

where $I$ denotes the indicator function. Then the quantity

$$e_n(f) = \frac{1}{n} \left[ \frac{f(0) + f(1)}{2} + \sum_{j=1}^{n-1} f\left(\frac{j}{n}\right) \right] - \int_0^1 f(x) \, dx$$

decreases as $O(1/n^2)$.

Proof. Given $n$, let $[i_n/n, (i_n+1)/n]$ be the subinterval that includes $c$. Then

$$e_n(f) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{f\left(\frac{i}{n}\right) + f\left(\frac{i+1}{n}\right)}{2} - \int_0^1 f(x) \, dx$$

$$= \frac{1}{n} \left[ \frac{f\left(\frac{i}{n}\right) + f\left(\frac{i+1}{n}\right)}{2} \right]$$

$$- \left[ \frac{f\left(\frac{i}{n}\right)}{2} \left( c - \frac{i}{n} \right) + \frac{f(c)}{2} \left( \frac{i+1}{n} - c \right) \right]$$

$$= \left( c - \frac{i}{n} \right) \left[ \frac{f\left(\frac{i+1}{n}\right) - f(c)}{2} \right] + \left( \frac{i+1}{n} - c \right) \left[ \frac{f\left(\frac{i}{n}\right) - f(c)}{2} \right]$$

$$= \text{(continued next page)}$$
\[ e_n(f) = \frac{s_2 - s_1}{2n^2} (nc - \lfloor nc \rfloor)(1 - (nc - \lfloor nc \rfloor)) \]

which is zero or decreases as \( O(1/n^2) \).

Extension to a general continuous piecewise linear function

\[ f(x) = \sum_{i=1}^{k} I_{[c_i, c_{i+1})} \|s_i(x - c_i) + d_i\| \]

is direct.

To prove Theorem 6 one needs only set \( f(x) = P(x) \) for \( 0 \leq x \leq 1 \).
This paper extends earlier results in the area of variance reduction techniques applied to simulation on a computer. In particular, it views the antithetic sampling technique as a combination of rotation and reflection sampling on a circle. The covariance structures induced by the techniques separately and together are derived and conditions under which they are optimal sampling plans are described. Rates of convergence for the variance of the sample mean are given for bounded, continuous and discrete functions and for unbounded
continuous functions with special, although commonly encountered, structure.

The advantage of reflection (basic antithetic) sampling is greatest when a certain symmetry property holds. Rotation-reflection sampling is superior to rotation sampling alone for continuous functions. In the bounded continuous case, the results show that rotation-reflection sampling speeds convergence to the large sample convergence rate achievable with rotation sampling alone. For the discrete case, rotation sampling does as well with regard to convergence as rotation-reflection sampling does. However, analysis of the discrete case shows that a sample size \( n \) may be considerably better than another \( n' \) although \( n' > n \).