An Optimal Multistage Sequential Monte Carlo Method

M.M.H. Ragheb, C.W. Maynard, and R.W. Conn

August 1974

UWFDM-107

FUSION TECHNOLOGY INSTITUTE

UNIVERSITY OF WISCONSIN

MADISON WISCONSIN

An Optimal Multistage Sequential Monte Carlo Method

M.M.H. Ragheb, C.W. Maynard, and R.W. Conn

Fusion Technology Institute
University of Wisconsin
1500 Engineering Drive
Madison, WI 53706

http://fti.neep.wisc.edu

August 1974

UWFDM-107

AN OPTIMAL MULTISTAGE SEQUENTIAL MONTE-CARLO METHOD

by

Magdi M. H. Ragheb
Charles W. Maynard
Robert W. Conn

July, 1974

Nuclear Engineering Department
The University of Wisconsin
Madison, Wisconsin 53706

FDM 107

ABSTRACT

Two unconstrained optimal multistage sequential Monte-Carlo schemes, and a constrained one, are exposed. The exposed method is a product of synthesis of multistage sequential Monte-Carlo, and of regression two-stage Monte-Carlo. The result is a combination of the advantages of the two methods with the elimination of the necessity to evaluate complicated functions or integrals, or matrix multiplications, in multistage sequential Monte-Carlo, and a faster convergence.

The method lends itself to efficient practical application whenever a high correlation exists between the used estimators.

Contents

1. Introduction

In this paper, a multistage regression sequential method is suggested. The procedure can be considered as a synthesis of the multistage sequential correlated sampling method as introduced by Halton[1,2,3,4,6], generalizing the two-stage sequential scheme of Marshall[7], and of two-stage Regression Monte-Carlo Methods[5,6,7,8].

Halperin[8], was the first to suggest almost linearly-optimal combination of estimators for minimal variance. His idea for the case of two-estimators was further considered by Spanier and Gelbard[5], Spanier[10], and Macmillan[9], in relation to applications in transport theory. Hammersley and Handscomb[6], also considered it with antithetic variates estimators. However in all these attempts[5,6,8,9,10] the analysis appears to have been for a two-stage process, and a multistage process similar to that of Halton[1,2,3,4,6] has not been tried. The suggestion of a multistage sequential optimal Monte-Carlo scheme is our aim in this paper.

2. Multistage Sequential Monte-Carlo:

Following an analysis similar to that of Halton[2], suppose we need to evaluate by Monte-Carlo the definite integral:

$$\mu = \int_0^1 f(x)dx \tag{2.1}$$

we use the primary estimates:

$$F(\S_1),\ f(\S_2)\cdot\ \cdot\ \cdot\ \cdot\ \cdot f(\S_n),\ \text{on } \{0,1\}$$

to yield the secondary estimate:

$$F_n = \frac{\sum_{i=1}^{n} f(\S_i)}{n} \tag{2.2}$$

which is an unbiased estimate for μ.

The variance would be:

$$\text{var} = \int_0^1 \{f(x) - \mu\}^2 dx \tag{2.3}$$

For variance reduction we rather sample a sequence of functions:

$$f_1(\eta_1),\ f_2(\eta_2) \cdots f_n(\eta_n)$$

where

$$f_n(y) = e_n(y) + \frac{f(y)}{g_n(y)} \tag{2.4}$$

For the case of correlated sampling (third sequential scheme in Halton's work[2]), we put $g_n(y) = 1$, and write:

$$e(y) = \mu - f(y) + h(y) \tag{2.5}$$

that is, $e_n(y)$ is to be an approximation $e(y)$ to a constant minus $f(y)$, ($h(y)$ small), Thus:

$$f_n(y) = e(y) + f(y) = \mu + h(y) \tag{2.6}$$

and:

$$\text{mean } (e + f) = \int_0^1 \{e(y) + f(y)\}\, dy = \mu \tag{2.7}$$

and:

$$\text{var } (e + f) = \int_0^1 \{h(y)\}^2 dy \tag{2.8}$$

also:

$$\int_0^1 e(y)dy = \int_0^1 h(y)dy = 0 \tag{2.9}$$

If h(y) is small, so is var(e + f), and the correlation coefficient of e and f is nearly (-1), as appears from (2.5).

It looks as if it is possible to have zero variance estimators; however such a case implies a knowledge of μ, the answer we are seeking. But, if we take successive samples to improve our estimation, better subsequent estimates can be obtained.

At stage n of the process, values of $f(\eta_i)$ at (n - 1) points: $\eta_1, \eta_2, \ldots$ η_{n-1} can be used to construct a function $G_n(y)$, such that $G_n(\eta_i)$ approximates to $f(\eta_i)$ for i = 1, 2, $\cdots$ (n - 1), and such that $\int_0^t G_n(y)dy$ is known. Then using the estimator function:

$$f_n(\S_n) = f(\S_n) + e_n(\S_n) = f(\S_n) + \int_0^1 G_n(y)dy - G_n(\S_n) = \mu + h_n(\S_n) \qquad (2.10)$$

(here we used equation (2.5)), we get:

$$\text{mean } (f_n) = \mu$$

and

$$\text{var } (f_n) = \int_0^1 \{h_n(y)\}^2 dy \leq \int_0^1 \{f(y) - G_n(y)\}^2 dy$$

Thus if:

$$\sup_{0 \leq y \leq 1} |f(y) - G_n(y)| \to 0 \text{ as } n \to \infty, \text{ the process is convergent.}$$

Halton[1,2,3,4] has studied the application of the above idea to the solution of linear systems of equations. For the system of N algebraic equations:

$$\underset{\sim}{A}\,\bar{x} = \bar{b} \qquad (2.11)$$

where $\underset{\sim}{A}$ is symmetric and positive definite (if not, $\underset{\sim}{A}$ can be premultiplied by its transpose $\underset{\sim}{A}'$), let:

$$\underset{\sim}{H} = \underset{\sim}{I} - c\,\underset{\sim}{A}, \text{ and } \bar{a} = c\,\bar{b} \qquad (2.12)$$

where $\underset{\sim}{I}$ is the identity matrix, and c is chosen such that all eigenvalues of $\underset{\sim}{H}$ are of modulus less than unity. (If eigenvalues of $\underset{\sim}{A}$ are $\alpha_i \geq 0$ and if $\bar{\alpha} \geq \max_i \alpha_i$, we take $c = 2/\bar{\alpha}$.)

Substituting in (2.11) we get:

$$\bar{x} = \underset{\sim}{H}\,\bar{x} + \bar{a} \tag{2.13}$$

The solution of (2.13) can then be considered as the convergent Neumann Series:

$$\bar{x} = (\underset{\sim}{I} - \underset{\sim}{H})^{-1}\,\bar{a} = (\underset{\sim}{I} + \underset{\sim}{H} + \underset{\sim}{H}^2 + \cdot\cdot\cdot\cdot)\,\bar{a} \tag{2.14}$$

After a fair amount of work, one should have a rough estimate $\bar{X}$ for $\bar{x}$. Then if we use the correction:

$$\bar{y} = \bar{x} - \bar{X} \tag{2.15}$$

and:

$$\bar{d} = \bar{a} + \underset{\sim}{H}\,\bar{X} - \bar{X} \tag{2.16}$$

Then equation (2.13) is transformed into another, of similar form:

$$\bar{y} = \underset{\sim}{H}\,\bar{y} + \bar{d} \tag{2.17}$$

where the elements of $\bar{d}$ are considerably smaller than those of $\bar{a}$.

This process was applied using the Von-Neumann Ulam estimator (actually the Wasow process[11]), by taking an arbitrary initial estimate $\bar{Z}^0$, and subsequently, for each stage of the process (stage is equivalent to complete history), taking:

$$\bar{d}^{(s)} = \bar{a} + \underset{\sim}{H}\,\bar{Z}^{(s-1)} - \bar{Z}^{(s-1)} \tag{2.18}$$

and

$$\bar{Z}^{(s)} = \bar{Z}^{(s-1)} + \bar{Y}^{(s)} = \sum_{r=1}^{s} \bar{Y}^{(r)} \tag{2.19}$$

where $Y_i^{(s)}$, are at each stage unbiased estimates of the correction $\{x_i - Z_i^{(s-1)}\}$, and $Z_i^{(s)}$ is an unbiased estimator for x_i. The score at each stage is:

$$Y_{ni}^{(s)} = \sum_{m=0}^{n} \left[\left\{ \prod_{y=1}^{m} \frac{H_{i_r - 1,\, i_r}}{P^{(s)}_{i_{r-1},\, i_r}} \right\} \cdot d^{(s)}_{i_m} \right] \tag{2.20}$$

here $P_{r,s}$ denotes the probability of transition from index r to index s over the random walk.

Thus:

$$\text{mean } (Y_i^{(s)}) = x_i - Z_i^{(s-1)} \tag{2.21}$$

and:

$$\text{mean } (Z_i^{(s)}) = (x_i), \tag{2.22}$$

the answer we are seeking to equation (2.13).

3. Regression Monte-Carlo:

As reported by Hammersley and Handscomb[6], if we have several unknown estimates $\mu_1, \mu_2, \cdots \mu_p$ and a set of estimators $t_1, t_2 \cdots t_n$ $(n \geq p)$ such that:

$$E\{t_i\} = x_{i1}\mu_1 + x_{i2}\mu_2 + \cdots + x_{ip}\mu_p \quad (i = 1, 2 \cdots n) \tag{3.1}$$

The minimum variance unbiased linear estimator of $\bar{\mu} = \{\mu_1, \mu_2, \cdots \mu_p\}$ will be:

$$t^* = (\underset{\sim}{X}' \underset{\sim}{V}^{-1} \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{V}^{-1} \bar{t} \tag{3.2}$$

where $\underset{\sim}{X}$ is the n x p matrix (x_{ij}), and $\underset{\sim}{V}$ is the n x n variance-covariance matrix of the t_i's, and where $\bar{t} = \{t_1, t_2, \cdots t_n\}$.

If $\underset{\sim}{V}_o \approx \underset{\sim}{V}$ is a reasonable approximation to $\underset{\sim}{V}$, then t_o^*:

$$t_o^* = (\underset{\sim}{X}' \underset{\sim}{V}_o^{-1} \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{V}_o^{-1} \bar{t} \tag{3.3}$$

will have a very nearly minimum variance, particularly since first-order deviations in x in the neighborhood of a minimum of a function F(x) only cause second order variations in F(x). Thus if $\underset{\sim}{V}$ is unknown, we may replace it by an estimate $\underset{\sim}{V}_o$. The sampling variance-covariance matrix of t_o^* is:

$$\text{var } t_o^* = (\underset{\sim}{X}' \underset{\sim}{V}_o^{-1} \underset{\sim}{X})^{-1} \tag{3.4}$$

In practice, a two-stage process is used. In the first stage, N independent sets of estimates $t_1, t_2, \cdots, t_n$ denoted by: $t_{1k}, t_{2k}, \cdots, t_{nk}$ $(k = 1, 2, \cdots N)$ are calculated from which v_{ij} are estimated by:

$$v_{ijo} = \sum_{k=1}^{N} (t_{ik} - \langle t_i \rangle)(t_{jk} - \langle t_j \rangle)/(N - 1) \tag{3.5}$$

where:

$$\langle t_i \rangle = (\sum_{k=1}^{N} t_{ik})/N \tag{3.6}$$

and: $\langle\ \rangle$ denotes an average value.

In the second stage, the estimator t_o^*, from Equation (3.3), is now used for another sample to get an estimate of $\bar{\mu}$. Hammersley and Handscomb[6], consider the case of the antithetic variates estimators:

$$t_1 = \tfrac{1}{2}f(\S) + \tfrac{1}{2}f(1-\S) \tag{3.7}$$

$$t_2 = \tfrac{1}{4}f(\tfrac{1}{2}\S) + \tfrac{1}{4}f(\tfrac{1}{2} - \tfrac{1}{2}\S) + \tfrac{1}{4}f(\tfrac{1}{2} + \tfrac{1}{2}\S) + \tfrac{1}{4}f(1 - \tfrac{1}{2}\S) \tag{3.8}$$

for the estimation of a single quantity:

$$\theta = \int_0^1 f(x)dx, \text{ where } f(x) = \frac{e^x - 1}{e - 1} \tag{3.9}$$

and report a "remarkable" gain in precision from the comparatively small extra labor by using the estimator:

$$t_o^* = 1.3411\ t_2 - 0.3411\ t_1 \tag{3.10}$$

4. Optimal Multistage Sequential Monte-Carlo:

4.1 Introduction:

Advantages of regression Monte-Carlo are that: firstly little, if any, bias is introduced to the answer, and secondly, if correlation exists in a given situation, considerable error reduction can be obtained; but if this correlation does not exist, nothing is lost in applying them (save some extra labor). The main advantage of the sequential Monte-Carlo method, is that successive samples are used to improve our knowledge of the situation, and we should be able to reduce the variances of subsequent estimates by adjusting our estimator functions. However, here arises a major disadvantage, namely difficulties in the construction of the function $G_n(y)$, as well as evaluation of the integral $\int_0^1 G_n(y)dy$ in equation (2.10). Even in the case of linear systems, the main advantage of the method is shadowed by the necessity of numerous matrix multiplications, as appears from equation (2.18).

A still unpublished trial to evaluate the term $\underset{\sim}{H}\,\overline{Z}^{(s-1)}$ in equation (2.18) by statistical means was done[11], to avoid matrix multiplication. It appears possible to achieve a hybridization, and deduce an optimal multistage sequential Monte-Carlo scheme in which advantages from each of the above-mentioned methods are combined. A multistage sequential process using sample values could improve the estimates of V_o from stage to stage, leading, hopefully, to a faster convergent process, and a smaller variance answer.

Though a generalization is an easy matter, for ease of demonstration and for simpler mathematical manipulation let us consider the particular case of equation (3.1) with $p = 1$, $n = 2$, $\overline{x} = \{1,1\}$. Consequently:

$$\underset{\sim}{V_o} = \begin{pmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{pmatrix}, \quad \underset{\sim}{V_o}^{-1} = \frac{1}{v_{11}v_{22} - v_{12}^2}\begin{pmatrix} v_{22} & -v_{12} \\ -v_{21} & v_{11} \end{pmatrix}, \quad \overline{x}\,\underset{\sim}{V_o}^{-1}\,\overline{x} = \frac{v_{11} + v_{22} - 2v_{12}}{v_{11}v_{22} - v_{12}^2}$$

the optimal variance is:

$$\text{var } t_o^* = (\bar{x}^{-1} \underset{\sim}{V}_o^{-1} \bar{x})^{-1} = \frac{v_{11}v_{22} - v_{12}^2}{v_{11} + v_{22} - 2v_{12}} \tag{4.1.1}$$

and the optimal estimator will be:

$$t_o^* = (\bar{x}' \underset{\sim}{V}_o^{-1} \bar{x})^{-1} \bar{x}' \underset{\sim}{V}_o^{-1} \bar{t} = \frac{(v_{11}v_{22} - v_{12}^2)}{v_{11}+v_{22}-2v_{12}} \cdot \frac{1}{(v_{11}v_{22}-v_{12}^2)} (v_{22}-v_{12} \quad v_{11}-v_{12}) \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

$$\text{or: } t_o^* = \frac{(v_{22} - v_{12})t_1 + (v_{11} - v_{12})t_2}{v_{11} + v_{22} - 2v_{12}} \tag{4.1.2}$$

for all: $v_{11}v_{22} - v_{12}^2 \neq 0$

or, in another notation:

$$\sigma_1^2 \sigma_2^2 \neq \rho_{12}^2 \sigma_1^2 \sigma_2^2 \quad , \text{ i.e. } \rho^2 \neq 1 \tag{4.1.3}$$

where ρ_{12} is the linear correlation coefficient, and σ^2 is the variance.

The condition (4.1.3) must not give us much worry, since in practical computations, $\rho_{12}^2 \neq 1$, unless t_1 and t_2 are just identical.

Defining:

$$a = \frac{v_{22} - v_{12}}{v_{11} + v_{22} - 2v_{12}} \tag{4.1.4}$$

one can write equation (4.1.2) in the form:

$$t_o^* = a\, t_1 + (1 - a)t_2 \tag{4.1.5}$$

which can just be identified as the minimal variance estimator considered by both Spanier[11] and Macmillan[9]. We adopt equation (4.1.5) for subsequent analysis of our suggested multistage sequential optimal process.

It is clear that: $\text{mean } t_o^* = \mu_1$ (4.1.6)

and that its variance is an optimal one.

4.2 First Optimal Sequential Scheme:

As an initialization, use is made of the two unbiased estimators t_1, t_2. After a fair amount of work, use equation (4.14) to get an initial estimate for a:

$$a^o \ \varepsilon(t_1, t_2) \tag{4.2.1}$$

where the ε here means: "estimated by use of".

Use is made now of this value of a^o to obtain an improved optimal estimator t_o^1* by using equation (4.1.5), and also to estimate an improved value for a, a^1. The multistage optimal scheme would look as follows:

initialization: t_1, t_2

$$\begin{array}{ll}
 & , a^o\varepsilon(t_1, t_2) \\
t_o^1* = a^o t_1 + (1-a^o)t_2 & \\
 & , a^1\varepsilon(t_o^1*, t_2) \\
t_o^2* = a^1 t_o^1* + (1-a^1)t_2 & \\
 & , a^2\varepsilon(t_o^2*, t_2) \\
t_o^3* = a^2 t_o^2* + (1-a^2)t_2 & \\
\vdots & , a^3\varepsilon(t_o^3*, t_2) \\
\cdots\cdots\cdots & \cdots\cdots\cdots \\
\cdots\cdots\cdots & , a^{n-2}\varepsilon(t_o^{(n-2)}*, t_2) \\
t^{(n-1)}* = a^{(n-2)} t_o^{(n-2)}* + (1-a^{(n-2)})t_2 & \\
 & , a^{n-1}\varepsilon(t_o^{(n-1)}*, t_2) \\
t^n* = a^{(n-1)} \ t_o^{(n-1)}* + (1-a^{(n-1)})t_2 &
\end{array} \tag{4.2.2}$$

Clearly at any stage, we have an optimally linear estimator, and at the n^{th} stage it can be easily proven that the stage optimal estimator obtained is an unbiased estimate** for the quantity of interest, so that it can be stopped after some

**Actually as pointed out by Hammersley and Handscomb[6], a very slight bias might be introduced by using sample values for $\underset{\sim}{V}_0$, rather than the exact values; but this does not limit the applicability of Regression methods.

error criterion is reached; the method thus lends itself easily to automatic computation. As an example at the 3rd stage:

$$E[t_o^{3*}] = E\{a^2 t_o^{2*} + (1-a^2)t_2\}$$

$$= E\{a^2[a^1 t_o^{1*} + (1-a^1)t_2] + (1-a^2)t_2\}$$

$$= E\{a^2[a^1\{a^o t_1 + (1-a^o)t_2\} + (1-a^1)t_2] + (1-a^2)t_2\} = \mu_1 \qquad (4.2.3)$$

and obviously, at each stage we are using a minimal variance estimator.

A disadvantage of that scheme is that after the initial stage, correlation between estimators is lost. Using the latest two best optimal estimators and combining them in an optimal matter would preserve correlation and speed convergence since our scheme here can be considered a case of correlated sampling.

4.3. Second Optimal Sequential Scheme:

Instead of using the best estimate from the last stage $t_o^{(n-1)*}$ to replace $t_o^{(n-2)*}$, and combine it with t_2, to get the optimal estimator t_o^{n*}, for the next stage, one could use the latest two (or more, for a generalized case) best values of t_o^*, to estimate the next stage optimal estimator, as in the following scheme:

initialization: t_1, t_2

$$t_o^{1*} = a^o t_1 + (1-a^o)t_2 \qquad a^o \varepsilon (t_1, t_2)$$

$$t_o^{2*} = a^1 t_o^{1*} + (1-a^1)t_2 \qquad a^1 \varepsilon (t_o^{1*}, t_2) \qquad (4.3.1)$$

$$t_o^{3*} = a^2 t_o^{2*} + (1-a^2)t_o^{1*} \qquad a^2 \varepsilon (t_o^{3*}, t_o^{1*})$$

$$
\begin{array}{ll}
 & a^3\varepsilon(t_o^{3*},t_o^{2*}) \\
t_o^{4*} = a^3 t_o^{3*} + (1-a^3)t_o^{2*} & \\
\vdots \quad \cdots\cdots\cdots & a^4\varepsilon(t_o^{4*},t_o^{3*}) \\
\vdots \quad \cdots\cdots\cdots & \cdots\cdots\cdots \\
t_o^{(n-1)*} = a^{(n-2)} t_o^{(n-2)*} + (1-a^{(n-2)})t_o^{(n-3)*} & a^{n-2}\varepsilon(t_o^{(n-2)*},t_o^{(n-3)*}) \\
 & a^{n-1}\varepsilon(t_o^{(n-1)*},t_o^{(n-2)*}) \\
t_o^{n*} = a^{(n-1)} t_o^{(n-1)*} + (1-a^{(n-1)})t_o^{n-2*} &
\end{array}
\qquad (4.3.1)
$$

As for the last process; $E(t_o^{n*}) = \mu_1$, which can be easily shown for the case n = 3:

$$E(t_o^{3*}) = E[a^2 t_o^{2*} + (1-a^2)t_o^{1*}] = E\{a^2[a^1 t_o^{1*} + (1-a^1)t_2] + (1-a^2)[a^o t_1 + (1-a^o)t_2]\}$$

$$= E\{a^2[a^1\{a^o t_1+(1-a^o)t_2\}+ (1-a^1)t_2] + (1-a^2)[a^o t_1+(1-a^o)t_2]\} = \mu_1 \qquad (4.3.2)$$

and clearly, at each stage, improved optimal estimators are used. In that scheme, t^{n*} and $t^{(n-1)*}$ can be expected to be highly correlated.

To avoid a possible loss in correlation on the first two stages, an alternate version of that scheme would be set up by using three originally highly correlated unbiased estimators for the same quantity μ_1 and following the scheme:

initialization: t_1, t_2, t_3

$$
\begin{array}{ll}
 & a^o\varepsilon(t_1,t_2),\ b^o\varepsilon(t_1,t_3) \\
\left[\begin{array}{l} t_o^{1*} = a^o t_1 + (1-a^o)t_2 \\ t_o^{2*} = b^o t_1 + (1-b^o)t_3 \end{array}\right. & \\
 & a^1\varepsilon(t_o^{1*},t_o^{2*}) \\
t_o^{3*} = a^1 t_o^{2*} + (1-a^1)t_o^{1*} & \\
 & a^2\varepsilon(t_o^{2*},t_o^{3*}) \\
t_o^{4*} = a^2 t^{3*} + (1-a^2)t^{2*} & \\
\vdots \quad \cdots\cdots\cdots & \cdots\cdots\cdots \\
 & a^{n-2}\varepsilon(t_o^{(n-1)*},t_o^{(n-2)*}) \\
t_o^{n*} = a^{(n-2)} t_o^{(n-1)*} + (1-a^{(n-2)})t_o^{(n-2)*} &
\end{array}
\qquad (4.3.3)
$$

4.4 Third Optimal Sequential Scheme:

In the first and second schemes t_1 and t_2 are assumed to be both unbiased estimators for the quantity of interest μ_1. Noting that a better variance reduction is obtained if the estimators t_1 and t_2 are highly correlated, one can consider the idea of considering t_2 as a highly correlated biased estimator for μ_1, and imposing a constraint that the expectation of the optimal estimator should be equal to μ_1.

Considering the estimator:

$$t^* = a + bt_1 + ct_2 \tag{4.4.1}$$

Let us impose first the constraint that:

$$E(t^*) = \mu_1 \tag{4.4.2}$$

which leads to the equation:

$$a + b\mu_1 + c\mu_2 = \mu_1 \tag{4.4.3}$$

where the $E(t_2)$ is equal to the biased value μ_2 for μ_1.

Now:

$$V(t^*) = \text{variance}(t^*) = E[t^* - E(t^*)]^2 = E[a+bt_1+ct_2-\mu_1]^2$$

$$= E[b(t_1-\mu_1) + c(t_2-\mu_2) + a - \mu_1 + b\mu_1 + c\mu_2]^2$$

$$= b^2\sigma_1^2 + c^2\sigma_2^2 + 2bc\ \text{cov}(1,2) + (a-\mu_1+b\mu_1+c\mu_2)^2$$

$$= b^2\sigma_1^2 + c^2\sigma_2^2 + 2bc\ \text{cov}(1,2) \tag{4.4.4}$$

by use of the constraint (4.4.3).

For a calculus relative minimum for the variance:

$$\frac{\partial V}{\partial b} = 0 \qquad b\sigma_1^2 + c\ \text{cov}(1,2) = 0$$

$$\frac{\partial V}{\partial c} = 0 \qquad b\ \text{cov}(1,2) + c\ \sigma_2^2 = 0$$

$$\text{or:} \begin{pmatrix} \sigma_1^2 & \text{cov}(1,2) \\ \text{cov}(1,2) & \sigma_2^2 \end{pmatrix} \begin{pmatrix} b \\ c \end{pmatrix} = 0 \tag{4.4.5}$$

For a non-trivial solution:

$$\sigma_1^2 \sigma_2^2 - [\text{cov}(1,2)]^2 = 0$$

or:

$$\sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 \rho_{1,2}^2 \quad , \sigma_1, \sigma_2 \neq 0$$

i.e.:

$$\rho_{1,2} = \pm 1 \tag{4.4.6}$$

Since, in practice, such a condition can not be met, we can not use equation (4.4.5), so we must content ourselves by using only the equation:

$$\frac{\partial V}{\partial c} = 0 \Rightarrow c = \frac{-b \text{ cov}(1,2)}{\sigma_2^2} \quad , b>0 \tag{4.4.7}$$

i.e. by solving for c in terms of b.

Substituting from (4.4.7) into (4.4.4), we get an expression for the minimum variance:

$$V_{min} = b^2 \sigma_1^2 (1 - \rho_{1,2}^2) \tag{4.4.8}$$

which, may be compared to equation (4.1.1), rewritten as:

$$\text{var}(t_o^*) = \frac{v_{11}v_{22} - v_{12}^2}{v_{11}+v_{22}-2v_{12}} = \frac{\sigma_1^2\sigma_2^2(1 - \rho_{1,2}^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2} \tag{4.4.9}$$

and one can relate* b^2 in the constrained case to the factor:

$$\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2} \quad , \text{ in the unconstrained case.}$$

Clearly better variance reduction will be obtained for higher correlation between t_1 and t_2.

---

*Just "relate", they are not identical.

From equations (4.4.3) and (4.4.7), we get:

$$a = \mu_1 + b(\rho_{1,2} \frac{\sigma_1}{\sigma_2} \mu_2 - \mu_1) \tag{4.4.10}$$

Thus our optimal constrained estimator would be:

$$t^* = \mu_1 + b\{(t_1-\mu_1) - \rho_{1,2} \frac{\sigma_1}{\sigma_2} (t_2-\mu_2)\} \tag{4.4.11}$$

Clearly: $E(t^*) = \mu_1$, and its variance is a minimum.

Let us notice that t_2 is only required to yield a high correlation coefficient $\rho_{1,2}$ with t_1, and contrary to the first two optimal sequential schemes, should not obligatorily be an unbiased estimator for μ_1, so that one can use a biased estimator t_2 for μ_1, such as a perturbed system ($E(t_2) = \mu_2$), and still get an unbiased estimate for t_1.

Like before, the use of the optimal estimator t* implies an a priori knowledge of the first two moments for t_1 and t_2; a sequential scheme can override this diffulty.

For the constant b>0, a choice of unity was found to be satisfactory. For that particular case:

$$c = \frac{-\text{cov}(1,2)}{\sigma_2^2} = -\rho_{1,2} \frac{\sigma_1}{\sigma_2} \quad , (b = 1) \tag{4.4.12}$$

$$a = -c\mu_2$$

$$t^* = t_1 - c\mu_2 + ct_2 = t_1 + c(t_2-\mu_2) \tag{4.4.13}$$

$$\text{var}(t^*) = \sigma_1^2(1 - \rho_{1,2}^2) \tag{4.4.14}$$

Notice the hidden resemblance between equations (4.4.13) and (2.10) or (2.18), however one does not need here to evaluate complicated integrals or

perform matrix multiplications, thus our schemes here improve over Halton's sequential Monte-Carlo method. However a basic difference can be noticed between equations (4.4.13) and (4.1.5), in the constrained and unconstrained cases.

An optimal sequencial scheme using the estimator (4.4.13) could be as follows (using the same procedure as the first sequencial scheme):

<u>initialization</u> t_1, t_2

$$c^o\varepsilon(t_1, t_2)\ ,\ \mu_2^o\varepsilon(t_2)$$

$$t_o^{1*} = t_1 + c^o(t_2 - \mu_2^o)$$

$$c^1\varepsilon(t_o^{1*}, t_2)\ ,\ \mu_2^1\varepsilon(t_2 \text{ on stages } 0,1)$$

$$t_o^{2*} = t_o^{1*} + c^1(t_2 - \mu_2^1)$$

$$c^2\varepsilon(t_o^{2*}, t_2)\ ,\ \mu_2^2\varepsilon(t_2 \text{ on stages } 0,1,2)$$

$$t_o^{3*} = t_o^{2*} + c^2(t_2 - \mu_2^2) \qquad (4.4.15)$$

$$c^3\varepsilon(t_o^{3*}, t_2)\ ,\ \mu_2^3\varepsilon(t_2 \text{ on stages } 0,1,2,3)$$

$$t_o^{(n-1)*} = t^{(n-2)*} + c^{n-2}(t_2 - \mu_2^{(n-2)})$$

$$c^{n-1}\varepsilon(t_o^{(n-1)*}, t_2)\ ,\ \mu_2^{(n-1)}\varepsilon(t_2 \text{ on stages } 0,1,2 \ldots (n-1))$$

$$t_o^{n*} = t_o^{(n-1)*} + c^{n-1}(t_2 - \mu_2^{n-1})$$

The value of μ_2 at each stage may be computed over all previous stages. Again, at any stage, we have an optimally linear estimator, and: $E(t_o^{n*}) = \mu_1$, e.g. for the 3rd stage:

$$E[t^{3*}] = E[t^{3*} + c^2(t_2 - \mu_2^2)] = E[t + c^1(t_2 - \mu_2^1) + c^2(t_2 - \mu_2^2)]$$

$$= E[t_1 + c^o(t_2 - \mu_2^o) + c^1(t_2 - \mu_2^1) + c^2(t_2 - \mu_2^2)] = \mu_1 \qquad (4.4.16)$$

If t_2 (the possibly biased estimator) is also an unbiased estimator for μ_1, one can similarly apply the second sequential optimal scheme in its two versions to the t_o^* constrained estimator.

5. Computational Formulae:

Recursive formulae adaptable to automatic computation are easily deducible. For the first sequential scheme, where at each stage the previous stage optimal estimator is combined in an optimal manner with t_2 one obtains:

$$\text{For } n = 3,\ t_{oI}^{3*} = a^2 a^1 a^o \cdot t_1 + (1 - a^2 a^1 a^o) t_2 \tag{5.1}$$

and for the nth stage:

$$t_{oI}^{n*} = \{(\prod_{s=o}^{n-1} a^s) \cdot t_1\} + \{(1 - \prod_{s=0}^{n-1} a^s) \cdot t_2\} =$$

$$= a^{n-1} \cdot d^{n-2} \cdot t_1 + (1 - a^{(n-1)} d^{(n-2)}) \cdot t_2 \ , \ n \geq 2 \tag{5.2}$$

so that one needs only to adjust the coefficients of t_1, t_2 at each stage.

For the third sequential scheme (constrained case) one gets:

$$\text{For } t = 3,\ t_{III}^{3*} = t_1 + (c^o + c^1 + c^2) t_2 - (c^o \mu_2^o + c^1 \mu_2^1 + c^2 \mu_2^2) \tag{5.3}$$

and for the nth stage:

$$t_{III}^{n*} = t_1 + (\sum_{s=o}^{n-1} c^s) \cdot t_2 - \sum_{s=o}^{n-1} (c^s \cdot \mu_2^s) =$$

$$= t_1 + (c^{n-1} + e^{n-2}) \cdot t_2 - [(c^{n-1} \mu_2^{n-1}) + (e^{n-2} \mu_2^{n-2})] \ , \ n \geq 2 \tag{5.4}$$

where: $d^n = \prod_{s=0}^{n} a^s$, and $e^n = \sum_{s=0}^{n} c^s$

A more elaborate recursive formula can be similarly deduced for the second multistage sequential optimal scheme.

6. Sample Problem:

We consider the example outlined by Hammersley and Handscomb[6], in relation to a two-stage regression Monte-Carlo process, and solve it by our suggested multistage sequential Monte-Carlo method. The problem is summarized in Eqns. (3.6), (3.7), (3.8) and (3.9), and use is made of the first multistage optimal sequential scheme. Some results are shown in Table (I). The data show that by a three stage process, corresponding to 150 histories by the optimal multistage sequential method, one can

Table I

	First multistage optimal sequential scheme = 0.418 027 3 (exact) N = # of samples/stage= = 50	Two-stage regression Monte-Carlo		
		N = 50	N = 200	N = 600
Stage I Initialization	$\bar{t}_1$ = 0.4216379 ± 0.0055377 $\bar{t}_2$ = 0.4189385 ± 0.0011407	0.4216379 ± 0.0055377 0.4189385 ± 0.0011407	0.4154351 ± 0.0026827 0.4173588 ± 0.0006926	0.4149366 ± 0.0014400 0.4172374 ± 0.0003676
Stage II	$\bar{t}_o^{1*}$ = 0.4180166 ± 0.0000228 $\bar{t}_2$ = 0.4168668 ± 0.0013059	0.4180166 ± 0.0000228 -----	0.4180305 ± 0.0000220 -----	0.4180219 ± 0.0000210 -----
Stage III	$\bar{t}_o^{2*}$ = 0.4180234 ± 0.0000161 $\bar{t}_2$ = 0.4195431 ± 0.0013390	-----	-----	-----
Stage VI	$\bar{t}_o^{6*}$ = 0.4180208 ± 0.0000172 $\bar{t}_2$ = 0.4183333 ± 0.0012358	-----	-----	-----
Batch Average over 9 stages	$\bar{t}$ = 0.4180175 ± 0.0000152	-----		

obtain better variance than that obtained by 400 or 1200 histories for the two-stage regression method, which could mean that the extra labor involved in the computation is counterbalanced by the variance reduction. These results are for the first scheme, where high correlation between the estimators is not preserved after the first few stages, however in the second sequential scheme this correlation might be preserved, and better results are likely to occur. There appears to be an optimal number of stages after which sampling should be stopped. These points are the subject of current research. A computer routine for that sample problem is listed in Appendix (I).

7. Conclusions:

The optimal multistage sequential Monte-Carlo method in its three versions combines the advantages of multistage sequential Monte-Carlo and of two-stage regression Monte Carlo: the need to evaluate difficult functions and integrals or to perform costly matrix multiplications is eliminated, with use of the information from successive samples to improve our estimation. A drastic decrease in the variance may be obtained whenever a high correlation exists between the used estimators.

Much work on the theory and on the practical applications of the approach remains to be done, especially for the derivation of highly correlated estimators. However, the present work was intended only for presentation and establishment of the idea of optimal multistage sequential Monte-Carlo.

Acknowledgement

Thanks are due to Professor J. H. Halton, Computer Sciences Department, The University of Wisconsin, for valuable discussions over his sequential Monte-Carlo method.

References

1. J. H. Halton, "Sequential Monte-Carlo," Proc. Camb. Phil. Soc., 58, p. 57 (1962).

2. J. H. Halton, "Sequential Monte-Carlo (Revised)," Mathematics Research Center, United States Army, The University of Wisconsin, Technical Report #816, (1967).

3. J. H. Halton, "A Retrospective and Prospective Survey of the Monte-Carlo Method," Computer Sciences Report #13, The University of Wisconsin, (1968).

4. J. H. Halton, "A Retrospective and Prospective Survey of the Monte-Carlo Method," SIAM Review, Vol. 12, No. 1, January (1970).

5. J. Spanier and E. M. Gelbard, "Monte-Carlo Principles and Neutron Transport Problems," Addison-Wesley Pub. Col., Reading, Mass. (1969).

6. J. M. Hammersley and D. C. Handscomb, "Monte-Carlo Methods," Methuen and Co., Ltd., p. 66, (1967).

7. A. W. Marshall, "The Use of Multistage Sampling Schemes in Monte-Carlo Methods," University of Florida Symposium on Monte-Carlo Methods. ed. H.A. Meyer, 37-46. New York; Wiley (1956).

8. M. Halperin, "Almost Linearly-Optimum Combination of Unbiased Estimates," J. Am. Stat. Association, 56, 36, (1961).

9. D. B. Macmillan, "Comparison of Statistical Estimators for Transport Problems," Nucl. Sci. Eng., 26, 366 (1966).

10. J. Spanier, "Two Pairs of Families of Estimators for Transport Problems," J. SIAM, Appl. Math. Vol. 14, No. 4, p. 702, July (1966).

11. J. H. Halton, personal communication.

Appendix (I): Computer Routine Listing for Sample Problem

```
                    FIRST MULTISTAGE OPTIMAL SEQUENTIAL SCHEME

      T1(X,C1)=C1*(EXP(X)+EXP(1.0-X)-2.0)
      T2(X,C2)=C2*(EXP(0.5*X)+EXP(0.5-0.5*X)+EXP(0.5+0.5*X)+EXP(1.0-0.5*
     XX)-4.0)
      TOSTR(X,C1,C2,A)=(A*C1*(EXP(X)+EXP(1.0-X)-2.0))+((1.0-A)*C2*(EXP (
     X0.5*X)+EXP(0.5-0.5*X)+EXP(0.5+0.5*X)+EXP(1.0-0.5*X)-4.0))
      ST1=0.0
      ST2=0.0
      STOSTR=0.0
      SSQ1=0.0
      SSQ2=0.0
      SQTO=0.0
      SR012=0.0
   99 READ 1,N,NN,NST
      XN=N
      XNN=NN
      XNST=NST
      XXNST=XNST-1.0
      D=SQRT(XN)
      DD=SQRT(XNN)
      CONST=1.0/(EXP(1.0)-1.0)
      C1=0.5*CONST
      C2=0.25*CONST
      K=1
      PRINT 1,K
      DO 70 I=1,N
      R=RANUN(RRR)
      F1=T1(R,C1)
      F2=T2(R,C2)
      ST1=ST1+F1
      ST2=ST2+F2
      SSQ1=SSQ1+F1*F1
      SSQ2=SSQ2+F2*F2
      SR012=SR012+F1*F2
   70 CONTINUE
      AV1=ST1/XN
      AV2=ST2/XN
      S1=((SSQ1/XN)-(AV1*AV1))
      S2=((SSQ2/XN)-(AV2*AV2))
      SG1= SQRT(S1)/D
      SG2= SQRT(S2)/D
      COV12=(SR012/XN)-(AV1*AV2)
      A=(S2-COV12)/(S1+S2-2.0*COV12)
      R012=(COV12/(SG1*SG2))/XN
      PRINT 2,N,AV1,AV2,SG1,SG2,S1,S2,COV12,R012,A
      ST2=0.0
      STOSTR=0.0
```

```
      SSQ2=0.0
      SQTO=0.0
      SRO12=0.0
      BAT=0.0
      BSQ=0.0
      DO 50 K=2,NST
      PRINT 1,K
      DO 60 J=1,NN
      R=RANUN(RRR)
      F3=TOSTR(R,C1,C2,A)
      F2=T2(R,C2)
      STOSTR=STOSTR+F3
      ST2=ST2+F2
      SQTO=SQTO+F3*F3
      SSQ2=SSQ2+F2*F2
      SRO12=SRO12+F2*F3
   60 CONTINUE
      AV2=ST2/XNN
      AV3=STOSTR/XNN
      S2=((SSQ2/XNN)-(AV2*AV2))
      S3=((SQTO/XNN)-(AV3*AV3))
      SG2=SQRT(S2)/DD
      SG3=SQRT(S3)/DD
      COV12=(SRO12/XNN)-(AV3*AV2)
      SO=(S2*S3-COV12*COV12)/(S2+S3-2.0*COV12)
      SON=SQRT(SO/XNN)
      AA=(S2-COV12)/(S3+S2-2.0*COV12)
      A=A*AA
      RO12=(COV12/(SG3*SG2))/XNN
      PRINT 2,NN,AV3,AV2,SG3,SG2,S3,S2,COV12,RO12,A,SO,SON
      BAT=BAT+AV3
      BSQ=BSQ+AV3*AV3
      ST2=0.0
      STOSTR=0.0
      SSQ2=0.0
      SQTO=0.0
      SRO12=0.0
   50 CONTINUE
      ST1=0.0
      SSQ1=0.0
      AVBAT=BAT/XXNST
      SIGB=SQRT(((BSQ/XXNST)-(AVBAT*AVBAT))/(XXNST-1.0))
      PRINT 4,AVBAT,SIGB
      GO TO 99
    1 FORMAT(3I5)
    2 FORMAT(I5/2E14.8/,2E14.8/,5E14.8,10X,2E14.8///)
    3 FORMAT(I5,3E14.8////)
    4 FORMAT(10X,16HBATCH STATISTICS,2(5X,E14.8)///)
      END
```