ARTICLE NO. 80-4044

Nonlinear Adaptive Estimation Procedures for Increasing the Efficiency in Monte Carlo Computations

Frederick H. Maltz* and Donald L. Hitzl†
Lockheed Research Laboratory, Palo Alto, Calif.

Estimator expansions in terms of orthonormal Hermite polynomials show particular promise for variance reduction in Monte Carlo computer simulations. By properly applying this approach, the Monte Carlo simulation will exhibit an acceleration toward convergence which, moreover, can be controlled by initial choices of certain parameters. Even further gains can be obtained by using adaptive procedures during the computations. In particular, an algorithm has been developed in which only those polynomial coefficients determined with sufficient precision are retained in the final Monte Carlo estimate. The underlying theory is briefly discussed and computational results are presented for several one-parameter model problems.

Introduction

THE original method of Chorin[1] for stochastic error variance reduction in Monte Carlo computations (Monte Carlo evaluation of multidimensional integrals) and extensions by Maltz and Hitzl[2,3] has recently been further improved as shown both in theory and in numerical experiments. Namely, by using certain adaptive estimation procedures to simultaneously identify the order of the nonlinearity and correct for it, additional gains have been demonstrated.

In this paper, we first briefly review the theory of these increased efficiency Monte Carlo estimators and then show how the variance in the resulting estimates can be reduced even further by using an adaptive series selection technique at each stage of the Monte Carlo computation. This new adaptive mechanization is illustrated here using some simple one-parameter examples. It should be emphasized, however, that these advanced Monte Carlo estimators were originally developed for application to multi-parameter trajectory simulations. In particular, estimates of increased precision were desired for such quantities as range and impact point which had been obtained previously by ordinary Monte Carlo averaging. A future publication[4] will show the gains that can be achieved when these adaptive estimators are applied to actual multi-parameter boost and reentry trajectory computations. Presently, we have computational experience with simulations involving 4, 5, 6, 8, 12, 17, 32, 35 and 40 random input parameters.

The idea of using stochastic series to achieve variance reduction in Monte Carlo computations is not new (see, for example, Ref. 5, pp. 69-73 where the orthonormal function method of Ermakov and Zolotukhin is described). The novelty of the Chorin technique lies in the fact that it is a straightforward, two-step procedure which can be readily implemented on the computer. Further, the basic Chorin estimator has desirable convergence properties for large sample sizes and the exact amount of variance reduction can readily be computed. In Refs. 2 and 3, the basic Chorin technique has been extended to multi-parameter problems and modified estimators of Chorin type have been developed which are more computationally efficient than the original Chorin estimator. This earlier work did not include the adaptive series selection option but, nevertheless, surprising reductions in Monte Carlo error variance were obtained for two model problems with 5 and 12 parameters, respectively.

The more general procedures we are currently using remove a basic deficiency in the original Chorin estimator. That is, the Monte Carlo sample values are utilized more effectively to obtain the final estimate. Even without the adaptive option these advanced estimators all have the characteristic that, after a certain number of Monte Carlo sample values, variance reduction is assured. The number N of individual Monte Carlos at which this first occurs is referred to as the "crossover point" N_c. For corrector series of high order, the crossover point will occur late (large N) but the subsequent reduction in variance will be significant. On the other hand, for a low order corrector series, the crossover point will occur earlier (smaller N) but the asymptotic variance reduction will not be as great. The purpose then of adaptive series selection is to achieve maximal variance reduction at each sample size N in the Monte Carlo run. Here this is accomplished by estimating the accuracy of individual terms in the corrector series and retaining only those terms which give a net variance decrease. This order identification technique then gives the best form of the corrector series at each step of the computation.

The Original Chorin Estimator

To begin, the basic Chorin estimator will be briefly reviewed. This estimator is formed in two steps as follows. The first step is to obtain a sample approximation for the random part of the function being integrated. This is referred to as the *corrector series* since it is used to reduce the variability in the final estimate. As we shall see, this corrector series is a minimum variance approximation to the random part (zero mean part) of the function being integrated. In particular, once the order of the series is fixed, the coefficients in the resulting corrector series are estimated using direct Monte Carlo averaging. The second step in the formation of the basic Chorin estimator is to compute, using a second (independent) set of sample values, the direct Monte Carlo average after subtracting the corrector series determined using the first sample set. In this way, the estimator will have a net reduction in error variance for a corrector series of a suitable order and for a sufficiently large Monte Carlo sample size.

The underlying approximation theory we use is that of stochastic series expansions. In particular, we use the or-

Received Dec. 18, 1978; revision received June 11, 1979.

Index categories: Analytical and Numerical Methods; LV/M Simulation; Entry Vehicle Mission Studies and Flight Mechanics.

*Research Scientist, Communications Sciences Laboratory; presently at Naval Underwater Systems Center, New London, Conn.

†Research Scientist, Advanced Systems Studies Department. Associate Fellow AIAA.

thonormal polynomial function expansion

$$y=f(x)=\sum_{k=0}^{\infty} a_k \Phi_k(x) \tag{1}$$

where $\Phi_k(x)$ are the orthonormal Hermite polynomials

$$\Phi_0=1 \quad \Phi_1=x \quad \Phi_2=\frac{x^2-1}{\sqrt{2}} \quad \Phi_3=\frac{x^3-3x}{\sqrt{6}}\ldots \tag{2}$$

and x is a single normal random variable with zero mean and unit variance. (The extension to a Gauss vector $\vec{x}$ with p independent components is presented in Ref. 3; however, only the one-dimensional case will be considered here). We assume that the function $f(x)$ has finite second moment

$$E[y^2]=\int_{-\infty}^{\infty} f^2(x)p(x)\,\mathrm{d}x<\infty \tag{3}$$

where $p(x)$ is the Gaussian density function $(1/\sqrt{2\pi})\exp(-x^2/2)$.

The coefficients a_k in Eq. (1) are given by

$$a_k=E[y\Phi_k(x)] \tag{4}$$

and convergence is understood to be in the stochastic sense of mean square error. Namely, the truncated expansion

$$f_n(x)=\sum_{k=0}^{n} a_k\Phi_k(x) \tag{5}$$

has a remainder $R_n \triangleq f(x)-f_n(x)$ such that

$$E[R_n]=0$$

and

$$\lim_{n\to\infty} E[R_n^2]=0 \tag{6}$$

Thus, the function $f_n(x)$ is the nth degree polynomial approximation to the original function f possessing minimum mean square error. Also, the mean of $f(x)$ is

$$\Theta \triangleq E[y]=a_0 \tag{7}$$

so we can rewrite Eq. (1) in the approximate form

$$y\cong a_0+\sum_{k=1}^{n} a_k\Phi_k(x) \tag{8}$$

which is the minimum variance, unbiased, nth degree polynomial approximation to the random variable y.

Now we are ready to give a precise definition of the Chorin estimator Θ_C for the mean of $y=f(x)$. Suppose we take two independent sets of M samples of x

$$\{x_1,x_2,\ldots x_M\} \quad \text{and} \quad \{x_1',x_2',\ldots x_M'\}$$

so there are $N=2M$ total samples. First, estimate each of the coefficients $(a_1,a_2,\ldots a_n)$ from the first set of samples as follows:

$$a_k^*=\overline{f(x)\Phi_k(x)} \qquad (k=1,2,\ldots,n) \tag{9}$$

by ordinary (or Direct) Monte Carlo

$$\overline{f(x)\Phi_k(x)}=\frac{1}{M}\sum_{j=1}^{M} f(x_j)\Phi_k(x_j) \tag{10}$$

Then form the *corrector series*

$$\Delta f=\sum_{k=1}^{n} a_k^*\Phi_k(x) \tag{11}$$

Next, using the second set of sample values, compute the sample average of

$$\Theta_C=\overline{f(x')-\Delta f(x')} \tag{12}$$

where

$$\overline{f(x')-\Delta f(x')}=\frac{1}{M}\sum_{j=1}^{M}[f(x_j')-\Delta f(x_j')] \tag{13}$$

It is important to note that the Chorin estimator Θ_C basically yields a higher precision estimate for the mean of f by subtracting off the *expected* stochastic fluctuations in $f(x')$ using $\Delta f(x')$ given by Eq. (11). Furthermore, the corrector series $\Delta f(x')$ has zero mean by construction. It has been shown[3] that this Chorin estimator Θ_C is unbiased and has variance

$$\sigma_C^2=M^{-1}A+M^{-2}B \tag{14}$$

where

$$A=\sum_{k=n+1}^{\infty} a_k^2 \qquad a_k \triangleq E[y\Phi_k] \tag{15}$$

and

$$B=\sum_{k=1}^{n} b_k^2 \qquad b_k^2 \triangleq \mathrm{Var}[y\Phi_k] \tag{16}$$

Substituting $N=2M$, the Chorin Monte Carlo error variance is

$$\sigma_C^2=2A/N+4B/N^2 \tag{17}$$

For comparison, the Direct (or ordinary) Monte Carlo error variance is

$$\sigma_D^2=\sigma^2/N \tag{18}$$

Note, for the stochastic series approximation to $f(x)$ given by Eq. (1),

$$\mathrm{Var}[y] \triangleq \sigma^2=\sum_{k=1}^{\infty} a_k^2 \tag{19}$$

so Eq. (17) has the simple interpretation that A is the mean square remainder in the stochastic series

$$A=E[R_n^2]=\sigma^2-\sum_{k=1}^{n} a_k^2 \le \sigma^2 \tag{20}$$

and b_k^2 is the mean square error in a one-sample Monte Carlo estimate of the coefficient a_k. Notice that as $N\to\infty$, the asymptotic Chorin variance is $\sigma_C^2 \to 2A/N$. The factor of 2 results from using only half the total number of samples in the final estimate Θ_C.

Increased Efficiency Estimators

As shown in Ref. 3, it is possible to improve the original Chorin estimator by allocating more samples to the final estimate. Alternatively, one can symmetrize the estimate. We consider here two methods of symmetrizing and the error variances associated with each.

The first "symmetrized Chorin" (SC) estimate is

$$\Theta_{SC}=(\Theta_C+\Theta_C')/2 \tag{21}$$

where Θ_C is the basic Chorin estimator given by Eq. (12) and Θ'_C is identical to it except for reversing the roles of the two sample sets $\{x\}$ and $\{x'\}$. As it turns out, the two estimates are not uncorrelated[3] and the net variance is

$$\sigma^2_{SC} = A/N + 2(B+C)/N^2 \tag{22}$$

where A and B are as before and

$$C = \sum_{j,k=1}^{n} (E[f\Phi_j\Phi_k])^2$$

The asymptotic variance for this estimate is $\sigma^2_{SC} \to A/N$ so the factor of 2 has now been eliminated. This reduced asymptotic variance is obtained, however, at the expense of the higher order penalty term $2CN^{-2}$.

We have found that by partitioning all the Monte Carlo samples into three independent sets, $\{x\}$, $\{x'\}$, and $\{x''\}$, forming three basic Chorin estimates

$$\Theta_C = \Theta_C(x,x') \quad \Theta'_C = \Theta_C(x',x'') \quad \Theta''_C = \Theta_C(x'',x) \tag{23}$$

and averaging these, a "tri-symmetrized Chorin" (TSC) estimate

$$\Theta_{TSC} = (\Theta_C + \Theta'_C + \Theta''_C)/3 \tag{24}$$

can be obtained in which the three component estimates are mutually uncorrelated. The net variance for this estimate is found to be

$$\sigma^2_{TSC} = A/N + 3B/N^2 \tag{25}$$

which should be compared to Eq. (17) for the basic Chorin variance. It is seen that Θ_{TSC} has a smaller Monte Carlo error variance than Θ_C under all conditions.

Since all these estimates are unbiased,

$$E\Theta_C = E\Theta_{SC} = E\Theta_{TSC} = \Theta \tag{26}$$

then Θ_{TSC} is a universally better estimate than the original estimator Θ_C devised by Chorin.

There are also other modified "Chorin-type" estimators that we have investigated and analyzed. However, these are beyond the scope of the present paper.

Adaptive Series Selection

To obtain the best possible results in Monte Carlo computer simulations using these Chorin type estimators, it is important to include only those terms in the corrector series which give a net variance decrease. This choice of terms depends upon the Monte Carlo run size N but, for each value of N, there is an optimal form for the corrector series. The method we use for this is referred to as an adaptive series selection algorithm. This algorithm will be described here for the basic Chorin estimator Θ_C but it has also been implemented with the two improved Chorin-type estimators Θ_{SC} and Θ_{TSC}.

The Monte Carlo error variance σ^2_C for the basic Chorin estimator was given previously [see Eq. (14)] with $M=N/2$. Using the definitions of A and B, this can be written in the more suggestive form

$$M\sigma^2_C = \sigma^2 - \sum_{k=1}^{n} (a_k^2 - M^{-1}b_k^2) \tag{27}$$

Equation (27) then leads to the following notation

$$M\sigma^2_C = \sigma^2 - \sum_{k=1}^{n} \Delta_k \tag{28}$$

where

$$\Delta_k = a_k^2 - M^{-1}b_k^2 \tag{29}$$

The quantity Δ_k, which can be either positive or negative depending upon the relative size of the two terms in Eq. (29), is easily seen to represent the net change in the variance as a result of estimating the kth coefficient a_k in the corrector series

$$\sum_{k=1}^{n} a_k\Phi_k(x)$$

Thus, we refer to the quantity Δ_k as the *variance reduction discriminant*.

In fact, writing the expansion [Eq. (1)] in the form

$$y = f(x) = \Theta + \sum_{k=1}^{\infty} g_k \tag{30}$$

with

$$g_k = a_k\Phi_k(x) \tag{31}$$

gives the following interpretation for Δ_k: a_k^2 is the variance of the component g_k in the polynomial expansion of f and $M^{-1}b_k^2$ is the error variance for the M-sample estimate of the coefficient a_k.

The discriminant Δ_k is then the basis for the adaptive series selection algorithm we have implemented. Both a_k and b_k^2 are estimated using all the Monte Carlo samples and the kth term retained only if the estimate Δ_k^* is positive. We refer to the set $\{\Delta_k^*\}$ as the "variance reduction spectrum." For each k in this procedure, it is to be expected that there is a minimum sample size $N_0(k)$ such that, for $N > N_0(k)$, $\Delta_k^* \geq 0$. Alternatively, we conjecture that for each given value of N, there is a point k_0 such that, for $k \geq k_0$, $\Delta_k^* < 0$. Also, in the application of this adaptive algorithm, particular care should be paid to the value Θ for the mean since, in many problems with large mean $|\Theta|$, Δ_k tends to become negative. (This is easily remedied, however, by subtracting an approximate value for the mean from all the data and working with the residuals.).

Test Functions

A total of nine one-dimensional functions $f_i(x)$ were devised by the authors in order to examine the variance reduction properties of both the original Chorin estimator Θ_C and the extended estimators Θ_{SC} and Θ_{TSC}. Each of the nine functions contains a single parameter k. These test functions are

$$f_1(x) = k\,\mathrm{sgn}(x)[1 - e^{-|x|/k}]$$

which has a mean of zero and variance

$$\sigma^2 = k^2\{1 - 2e^{1/2k^2}[1 - \mathrm{erf}(1/\sqrt{2}k)] + e^{2/k^2}[1 - \mathrm{erf}(\sqrt{2}/k)]\}$$

$$f_2(x) = (2/\pi)\tan^{-1}kx$$

which again has a mean of zero and variance given by the definite integral

$$\sigma^2 = \left(\frac{2}{\pi}\right)^{5/2}\frac{1}{k}\int_0^{\pi/2} \Theta^2 \exp\left(-\frac{\tan^2\Theta}{2k^2}\right)\sec^2\Theta\, d\Theta$$

$$f_3(x) = k\,\mathrm{sgn}x$$

which has a mean of zero and variance $\sigma^2 = k^2$.

$$f_4(x) = e^{-kx^2}$$

which is an even function with mean $\Theta = 1/\sqrt{2k+1}$ and variance $\sigma^2 = 1/\sqrt{4k+1} - 1/(2k+1)$.

$$f_5(x) = \begin{cases} x & \text{for } |x| \leq k \\ k\,\mathrm{sgn}x & \text{for } |x| > k \end{cases}$$

which is a linear clip (or saturation) function with $\Theta = 0$ and variance

$$\sigma^2 = k^2 - \sqrt{2/\pi} k e^{-k^2/2} + (1-k^2)\operatorname{erf}(k/\sqrt{2})$$

Finally, the last four test functions are simply the first four orthonormal Hermite polynomials with an additive constant k for a nonzero mean.

$$f_6(x) = k + \Phi_1(x) = k + x$$

$$f_7(x) = k + \Phi_2(x) = k + (x^2 - 1)/\sqrt{2}$$

$$f_8(x) = k + \Phi_3(x) = k + (x^3 - 3x)/\sqrt{6}$$

and

$$f_9(x) = k + \Phi_4(x) = k + (x^4 - 6x^2 + 3)/\sqrt{24}$$

For each of these last four functions, the variance $\sigma^2 = 1$.

Computational Results

A digital computer program VARED (Variance Reduction) incorporating all three Chorin-type estimators was written in order to obtain some computational experience with these advanced Monte Carlo estimators. The computer program was run on the Univac 1110 at LMSC. Results are presented in Figs. 1-8 for the four test functions $f_4(x)$, $f_5(x)$, $f_6(x)$, and $f_7(x)$ with the single parameter k set equal to 1 in each case. With this choice for k, the exact mean values are

$$\Theta = 1/\sqrt{3} = 0.57735..., \quad 0, \quad 1 \quad \text{and} \quad 1 \quad \text{respectively}$$

while the variance are

$$\sigma_f^2 = 0.11388..., \quad 0.51606..., \quad 1.0 \quad \text{and} \quad 1.0$$

The first plot for each function (Figs. 1, 3, 5, 7) displays the computed mean value vs the number of function evaluations N for each of the four estimators Θ_D, Θ_C, Θ_{SC}, and Θ_{TSC}. Note the acceleration toward convergence realized in each case using the extended Chorin-type estimators. Also, the adaptive algorithm was employed for all the results presented here. In particular, the algorithm was able to distinguish successfully between even and odd functions so, for an odd function, all even coefficients $a_2, a_4, ...$ in the corrector series Eq. (11) were automatically zeroed by the computer.

The real payoff for these variance reduction methods is shown in the second plot for each function (Figs. 2, 4, 6, 8) where the Monte Carlo error variance is shown as a function of N. The plots can be interpreted in either of two ways. For a given N, the reduction in variance is immediately available while, for a given variance, the allowable reduction in sample size N can be easily assessed. Also, it is important to note the far superior performance of the "symmetrized" estimators

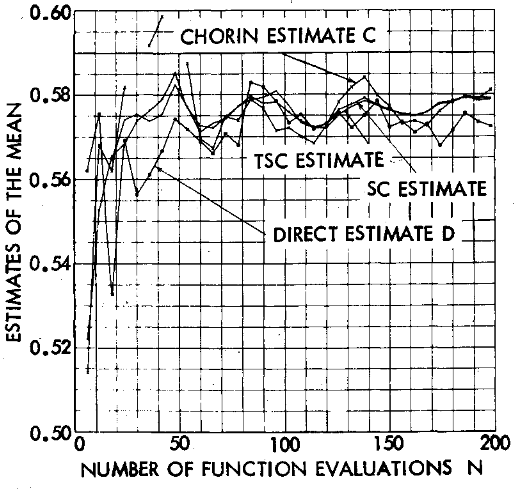

Fig. 1 Estimates of the mean of $f_4(x)$ as a function of N with $k=1$ and $n=4$. Results using all four estimators are shown. The exact mean value is $1/\sqrt{3} = 0.57735...$.

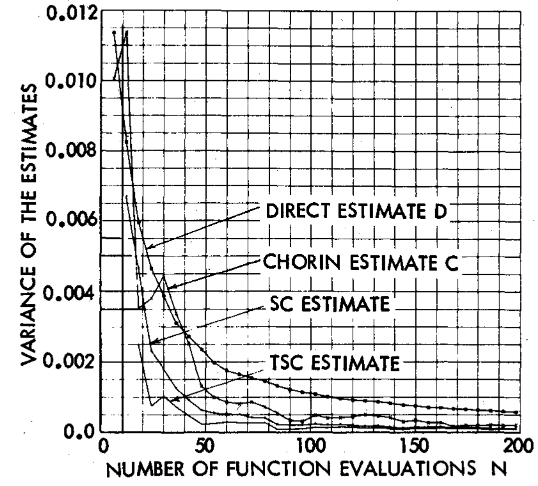

Fig. 2 Variances for the four estimates shown in Fig. 1. Note the final crossover around $N=40$ for the C estimate.

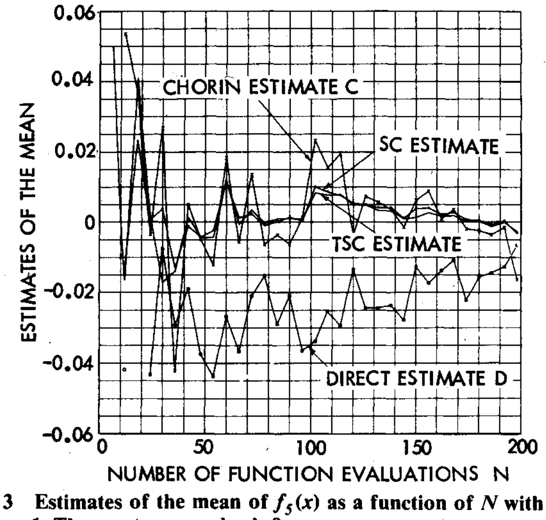

Fig. 3 Estimates of the mean of $f_5(x)$ as a function of N with $k=1$ and $n=1$. The exact mean value is 0.

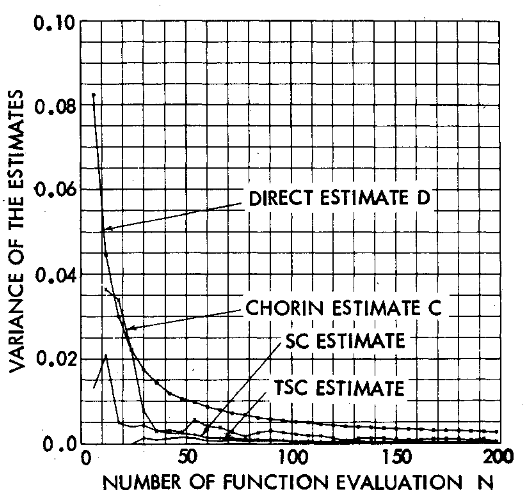

Fig. 4 Variances for the four estimates shown in Fig. 3.

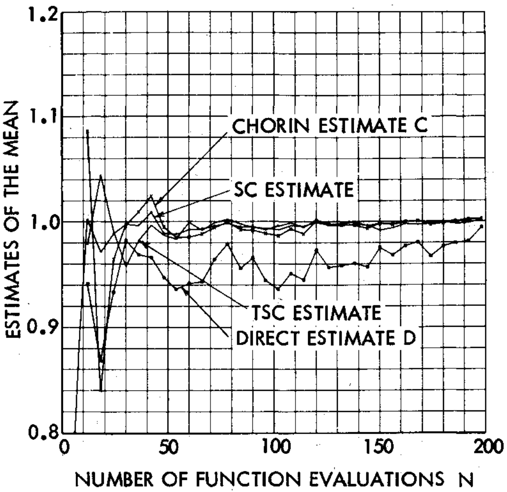

Fig. 5 Estimates of the mean of f_6 (x) as a function of N with $k=1$ and $n=1$. The exact mean value is $k=1$.

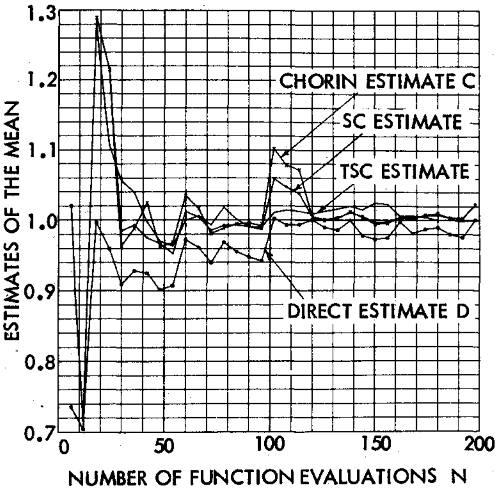

Fig. 7 Estimates of the mean of f_7 (x) as a function of N with $k=1$ and $n=2$. The exact mean value is $k=1$.

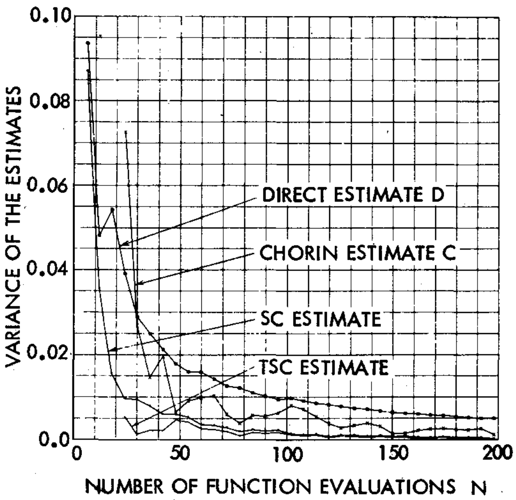

Fig. 6 Variances for the four estimates shown in Fig. 5.

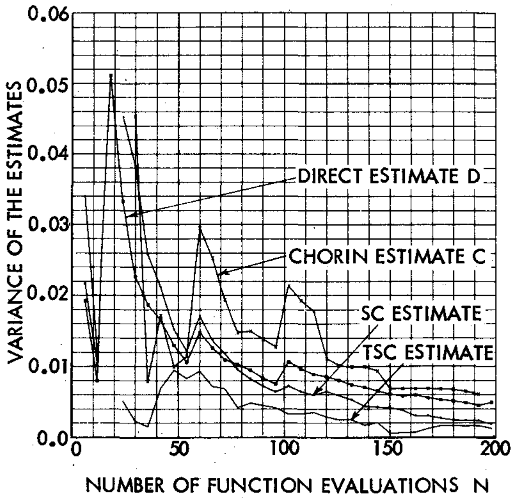

Fig. 8 Variances for the four estimates shown in Fig. 7.

Θ_{SC} and Θ_{TSC} in all cases compared to the direct estimator Θ_D and basic Chorin estimator Θ_C. This is due to the more effective use of the available samples in determining both the corrector series Δf and the final estimates Θ_{SC} and Θ_{TSC}.

To investigate this a little more closely, consider, for example, Figs. 5 and 6 for the function $f_6(x)=1+x$. For full correction in this case, $n=1$ so only the first term $a_1{}^*\Phi_1(x)$ in Eq. (11) is retained. Then, as N increases, ordinary Monte Carlo estimates of (slowly!) increasing precision are obtained for both the mean a_0 and the correction series coefficient a_1. (Note the exact values are $a_0=a_1=1$.) Consequently, we have $A\rightarrow 0$ from Eq. (20) and $B\rightarrow 3$ from Eq. (16) which gives the following theoretical limiting values

$$\sigma_D^2=1/N \qquad \sigma_C^2=12/N^2 \qquad \sigma_{TSC}^2=9/N^2 \tag{32}$$

from Eqs. (18), (17), and (25), respectively. Examining Fig. 6 shows that these limiting values are indeed approached by the computer simulation but that the estimated Chorin variance σ_C^2 is initially quite erratic and the final convergence is much slower than that obtained for the "symmetrized" variances σ_{SC}^2 and σ_{TSC}^2. This is caused, when N is small, by the relatively poor estimates obtained for a_1 using the original Chorin approach and *not* by estimating additional coefficients in the corrector series which, in fact, should not be included.

Conclusions

As shown in the computational results for the single parameter test functions, the adaptive series selection algorithm has been found to be an essential adjunct to the improved efficiency Monte Carlo estimators described here. The adaptive technique is based on the fact that the variance reducing (or increasing) effect of each term in the complete (and generally nonlinear) correction series can be isolated. This important result is obtained by inspection of the exact theoretical expression for the estimator variance and leads to the so-called variance reduction discriminant given by Eq. (29). In this equation, the first quantity a_k^2 determines *how important* a given term is in the corrector series (net reduction in the mean square remainder) while the second quantity $M^{-1}b_k^2$ determines *how accurately* the given term can be estimated (net reduction in the sum of the coefficient error variances).

In conclusion, since the adaptive algorithm is based on solid theoretical grounds and the computer simulations demonstrate that it indeed works, we believe this is an important development in Monte Carlo computations. Presently, we are unaware of any other similar work in the area of adaptive variance reduction techniques.

Acknowledgments

It is a sincere pleasure to thank F. Zele of Lockheed Missiles and Space Company for his help in developing and refining our Monte Carlo computer programs. Also, we are grateful to G. Morgan for her efforts in typing the manuscript.

References

[1]Chorin, A.J., "Hermite Expansions in Monte-Carlo Computation," *Journal of Computational Physics,* Vol. 8, Dec. 1971, pp. 472-482.

[2]Maltz, F.H., "Hermite Accelerated Monte-Carlo for Evaluating Multi-Target Tracking Filters," *Proceedings IEEE Conference on Decision and Control,* New Orleans, La. 1977.

[3]Maltz, F.H. and Hitzl, D.L., "Variance Reduction in Monte Carlo Computations Using Multi-Dimensional Hermite Polynomials," *Journal of Computational Physics,* Vol. 32, No. 3, Sept. 1979, pp. 345-376.

[4]Hitzl, D.L. et al., "Applications of an Adaptive Procedure to Accelerate Monte Carlo Computations," in preparation.

[5]Hammersley, J.M. and Handscomb, D.C., *Monte Carlo Methods,* Methuen and Co. Ltd., London, 1964.