MONTE CARLO VARIANCE REDUCTION METHODS

11/4/2013

1. INTRODUCTION

The question arises of whether one can reduce the variance associated with the sampling of a random variable? Indeed we can, but we need to be somehow sophisticated in sampling the random variable. There is a wealth of methods and techniques that allow us to reduce the variance: these are called "variance reduction methods." We shall consider some of them here. These methods are sometimes problem-dependent, but their usage in some cases would make possible the solution of problems that would be otherwise intractable, because of the tremendous amount of computer time that would be needed in a direct simulation. This is the cause of the misconception of the Monte Carlo method being "expensive." Indeed, if variance reduction methods are not used in some cases, the numerical experiment may become more costly than building a real physical experiment. The knowledge of variance reduction methods is thus a requirement for the clever and economical use of the Monte Carlo method.

2. VARIANCE REDUCTION BY MODIFICATION OF THE SAMPLING SCHEME

Every random variable has an underlying variance associated with its probability density function. For instance, the random variable:

ξ: Points obtained in throwing a single die

$$\xi : \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix} \tag{1}$$

The mean value and the variance of this random variable are:

$$\mu(\xi) = \frac{1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}}{\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}} = \frac{21}{6} = 3.5$$

$$\sigma^2(\xi) = \frac{1^2\times\frac{1}{6}+2^2\times\frac{1}{6}+3^2\times\frac{1}{6}+4^2\times\frac{1}{6}+5^2\times\frac{1}{6}+6^2\times\frac{1}{6}}{\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}} - (3.5)^2$$

$$= \frac{91}{6} - 12.25 = 2.917$$

Most variance reduction methods depend on simple, but ingenious modifications of the sampling scheme used in the estimation of a given random variable. A user must in general use physical or mathematical insight into the problem considered. The key idea is to use an altered probability density functions that have the *same mean*, but *lower variances*. To demonstrate this not so intuitive possibility of variance reduction, we consider this modification of the sampling scheme for the die throwing experiment as follows:

$$\xi = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1+6}{2} & \frac{2+5}{2} & \frac{3+4}{2} \end{pmatrix} \qquad (2)$$

The first column is the case in which an outcome of 1 or 6 is obtained, the second column the 2 or 5 outcomes, and the third column the 3 or 4 outcomes. Of course the probability of either one of these events is the sum of each probability of 1/6, which is now 1/3.

Let us first calculate the mean value of this modified sampling distribution:

$$\mu = \frac{1}{3}\left(\frac{1+6}{2}\right) + \frac{1}{3}\left(\frac{2+5}{2}\right) + \frac{1}{3}\left(\frac{3+4}{2}\right) = \frac{7}{2} = 3.5,$$

which is the same mean as obtained from the distribution of Eq. (1). However, let us now estimate the variance:

$$\begin{aligned}\sigma^2 &= \left[\frac{1}{3}\left(\frac{1+6}{2}\right)^2 + \frac{1}{3}\left(\frac{2+5}{2}\right)^2 + \frac{1}{3}\left(\frac{3+4}{2}\right)^2\right] - \left(\frac{7}{2}\right)^2 = \\ &= \left[\frac{1}{3}\left(\frac{7}{2}\right)^2 + \frac{1}{3}\left(\frac{7}{2}\right)^2 + \frac{1}{3}\left(\frac{7}{2}\right)^2\right] - \left(\frac{7}{2}\right)^2 \\ &= 0.0\end{aligned}$$

Here we are not just reducing the variance, but we are even obtaining a *zero*-variance scheme. Of course it is not always possible in practice to obtain a zero variance scheme, but the example shows that we surely can always modify a sampling scheme so as to preserve the value of the mean, but decrease the value of the variance. As shown later, what was done here is no magic, but was a subtle application of the Antithetic Variates variance reduction method.

Since we have already established that the error:

$$\varepsilon \propto_{\ell} \frac{\sigma}{\sqrt{N}},$$

it is clearly observed that to decrease the error through reducing the value of σ is a much more efficient process than reducing it just by increasing the number of trials in the denominator (N). It would require an increase of the number of samples to 4N, to reduce the error in the estimate to just a factor of ½.

Various techniques for variance reduction will be considered next.

3. ANALOG, HIT OR MISS, OR POOR MAN'S SAMPLING

We consider variance reduction for the estimation of simple integrals since the basic principles involved generalize quite easily to more complex problems, such as particle transport and fluid flow.

Suppose we need to estimate the integral:

$$\theta = \int_0^1 f(x)dx \tag{3}$$

with: $0 \le f(x) \le 1,$ when $0 \le x \le 1$.

Let us draw the curve:

$$y = f(x) \ ,$$

within the unit square, as shown in Fig. 1.

We may write:

$$f(x) = \frac{\int_0^1 g(x,y)dy}{1^2} \ , \tag{4}$$

where: $g(x)$ = 0, if $f(x) \le y$
= 1, if $f(x) > y$.

Simply stated, we accept any sampled points that lie below the curve for the mean estimation process and reject those that do not.

We then can estimate θ as a double integral:

$$\theta = \int_0^1 \int_0^1 g(x,y)dxdy,$$

using the estimator:

$$\bar{g} = \frac{1}{n}\sum_{i=1}^{n} g\left(\xi_{2i-1,}\xi_{2i}\right) = \frac{n^*}{n} \tag{5}$$

where: n^* is the number of occasions on which $\xi_{2i} \le f\left(\xi_{2i-1}\right)$,

ξ_{2i} is a uniformly sampled y-coordinate point over the interval [0,1],

ξ_{2i-1} is a uniformly sampled x-coordinate point over the unit interval,

$\left(\xi_{2i-1}, \xi_{2i}\right)$ thus determines a point uniformly distributed over the unit square.

In fact, we sample n points at random in the unit square, and count the proportion of them which lie below the curve $y = f(x)$. This is in retrospect the rejection sampling method.

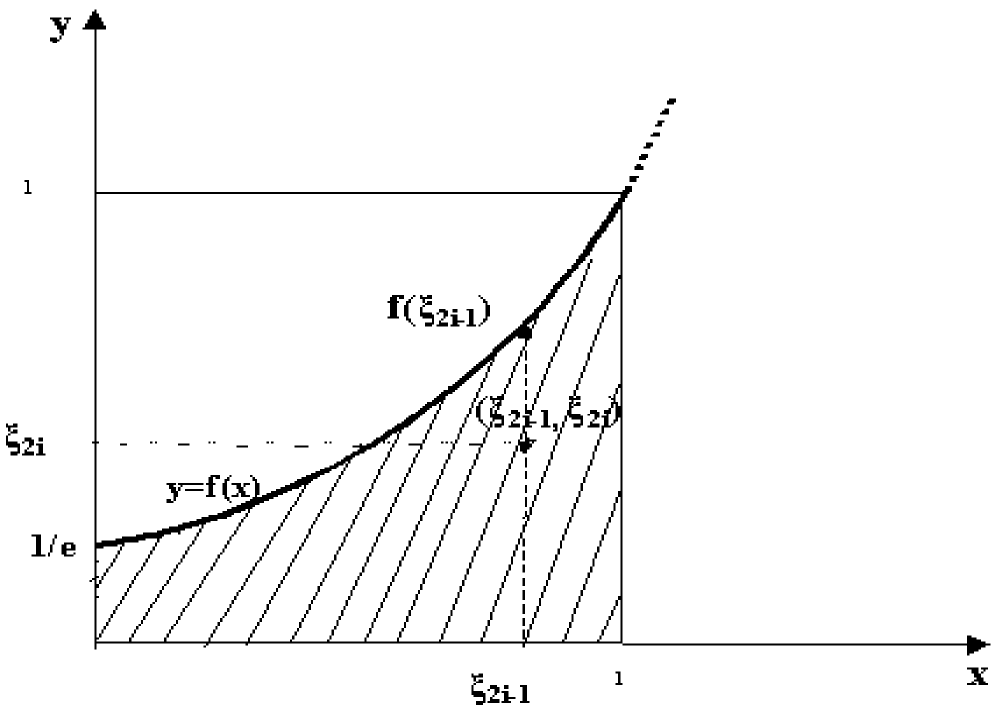

Figure 1. Hit-or-Miss, Analog, or Poor Man's Monte Carlo estimation for the estimation of an integral.

As a numerical example, let us estimate the integral:

$$I = \int_0^1 e^x dx \tag{6}$$

which has an exact analytical value:

$$I = e^x \Big|_0^1 = 2.718281 - 1.0 = 1.718281$$

To have $0 \le f(x) \le 1$ when $0 \le x \le 1$, we perform a simple scaling:

$$\theta = I = e\int_0^1 \frac{e^x}{e}dx = e\int_0^1 g(x)dx = e\theta' \tag{7}$$

and we then estimate the value of θ' from which we can evaluate θ. A procedure using the Analog Monte Carlo Integration method is shown in Fig. 2.

```
!	analog.f90
!	Analog, Hit or Miss, or Poor Man's Monte Carlo
!	Estimation of the value of integral using Analog Monte Carlo
!	I=Integral e**x over the interval [0,1] = 1.718281828
!	M. Ragheb
	program analog
	real mean
	real :: e=2.718281828
	real :: trials=1000
!	Define sampled function; function is scaled to lie within the
!	unit square
	g(x) = exp(x)/e
!	Initialize score
	score = 0.0
!	Sample points on unit square
	do i=1,trials
		call random(rr)
		x=rr
		f=g(x)
		call random(rr)
		y=rr
		if(y.LE.f)then
			score =score+1.0
		end if
	end do
!	Estimate mean value
	mean=score/trials
!	rescale the mean value
	mean = mean * e
!  Write results
	write(*,*) mean, trials
	end
```

Figure 2. Procedure for integration using Analog, Hit-or-miss, or Poor Man's Monte Carlo.

In the case of the hit and miss Monte Carlo, the variance is that of the binomial distribution given by:

$$\sigma^2 = p(1-p),$$
$$\sigma = \sqrt{p(1-p}$$

In our case, p is the number of hits to the number of trials ratio:

$$p = \frac{n^*}{n}$$

Some computational results are shown in Table 1.

As an example to estimate the variance and the standard deviation of the binomial distribution:

$$p = \frac{n^*}{n} = \frac{score}{trials} = \frac{1.716323}{2.718281828} = 0.631399946216$$
$$(1 - p) = 1 - 0.631399946216 = 0.368600053784$$
$$\sigma^2 = p(1-p) = 0.631399 \times 0.368600 = 0.232733$$
$$\sigma = \sqrt{p(1-p} = \sqrt{0.232733} = 0.482424$$

Table 1. Mean values and Fractional Standard Deviation using Analog, Hit-or-miss, or Poor Man's Monte Carlo.

$\theta \pm \frac{\sigma}{\sqrt{n}}$	$\theta \pm \sigma$	fsd $\left[\frac{\sigma / \sqrt{n}}{\theta}\right]$	n
1.716323± 0.003411	1.716323 ± 0.482388	0.001987	20,000
1.756010± 0.015129	1.756010 ± 0.478420	0.008616	1,000
1.658151± 0.049020	1.658151 ± 0.490200	0.029563	100

It can be noticed that an excessive number of trials is required to get a low fractional standard deviation. The standard deviation is that of the binomial distribution.

4. CRUDE MONTE CARLO SAMPLING

If ξ_1, ξ_2, ξ_3, …, ξ_n are independent random numbers uniformly distributed over the unit interval [0.1], then the quantities:

$$f_i = f(\xi_i) \tag{8}$$

are independent random variates with expectation θ. Therefore, according to the Central Limit Theorem:

$$\bar{f} = \frac{1}{n} \sum_{i=1}^{n} f_i \tag{9}$$

Table 2. Mean Values and Standard Deviations using Crude Monte Carlo.

$\theta \pm \frac{\sigma}{\sqrt{n}}$	$\theta \pm \sigma$	fsd $\left[\frac{\sigma / \sqrt{n}}{\theta}\right]$	n
1.720654 ± 0.003475	1.720654 ± 0.491439	0.002019	20,000
1.725170 ± 0.015334	1.725170 ± 0.484903	0.008888	1,000
1.772332 ± 0.049850	1.772332 ± 0.498500	0.028127	100

is an unbiased estimator of θ, and its variance is:

$$\frac{1}{n}\int_0^1 \left(f(x) - \overline{f} \right)^2 dx = \frac{\sigma^2}{n} \tag{10}$$

The standard error of $\bar{f}$ is:

$$\sigma_{\bar{f}} = \frac{\sigma}{\sqrt{n}}.$$

From Table 2, it appears that Crude Monte Carlo does not improve in term of the achieved error estimate over Hit or Miss Monte Carlo.

In Fig. 3 we show how simple the procedure for crude Monte Carlo for the estimation of the integral is. In Fig. 4 the procedure is modified for the estimation of the variance using the conventional estimation formula. Figure 5 shows an alternative way of estimating the variance using the earlier derived recursive variance estimation formula.

```
!       crude f90
!       Crude Monte Carlo
!       I=Integral[e**x, 0,1]=1.718281828
!       M. Ragheb
        program crude
        real mean
        real :: trials=100
!       Define sampled function
        g(x) = exp(x)
!       Initialize score
        score = 0.0
!       Sample function
        do i=1,trials
                call random(rr)
                score =score+g(rr)
        end do
!       Estimate mean value
        mean=score/trials
!  Write results
        write(*,*) mean, trials
        end
```

Figure 3. Procedure for integration using Crude Monte Carlo.

```
!       crude1.f90
!       Crude Monte Carlo
!       Variance estimation using conventional formula
!       I=Integral[e**x ,0,1]=1.718281828
!       M. Ragheb
        program crude1
        real mean
        real :: trials=20000
!       Define sampled function
        g(x) = exp(x)
!       Initialize scores
        score = 0.0
        sum_squares = 0.0
!       Sample function
        do i=1,trials
```

```
                call random(rr)
                s = g(rr)
                score =score + s
                sum_squares = sum_squares + s *s
        end do
!       Estimate mean value, variance, standard deviation,
!       and fractional standard deviation
        mean=score/trials
        variance = sum_squares/trials - mean*mean
        variance = (trials/(trials-1.0)) * variance
        std_dev = sqrt (variance)
        std_error = std_dev/sqrt(trials)
        frac_stan_dev = std_error/mean
!  Write results
        write(*,*) mean, std_error, frac_stan_dev, trials
        end
```

Figure 4. Procedure for integration using Crude Monte Carlo, with conventional formula for the estimation of the variance.

```
!       crude2.f90
!       Crude Monte Carlo
!       Variance estimation using recursive formula
!       I=Integral[e**x, 0,1]=1.718281828
!       M. Ragheb
        program crude2
        real mean_value, t, ta
        real :: trials=100
!       Define sampled function
        g(x) = exp(x)
!       Initialize variables
        t=0.0
        s_square_old = 0.0
!       Sample function
        do k=1,trials
                xk=k
                call random(rr)
                s = g(rr)
                if (k.GE.2) then
                        ta = t/(xk-1.0)
                        s_square_new =   (xk-2.0)/(xk-1.0)*s_square_old          &
&                                          + (ta-s)*(ta-s)/xk
                        s_square_old = s_square_new
                end if
                t=t+s
        end do
!       Estimate mean value, variance, standard deviation,
!       and fractional standard deviation
        mean_value = t/trials
        std_dev = sqrt (s_square_new)
        std_error = std_dev/sqrt(trials)
        frac_stan_dev = std_error/mean_value
!  Write results
        write(*,*) mean_value, std_error, frac_stan_dev, trials
        end
```

Figure 5. Procedure for integration using Crude Monte Carlo, with recursive formula for the estimation of the variance.

5. CORRELATED SAMPLING OR CONTROL VARIATES

The equation:

$$\theta = \int_0^1 f(x)dx$$

is broken into two parts by adding the subtraction an easy function $\phi(x)$ that approximates the function f(x):

$$\theta = \int_0^1 \phi(x)dx + \int_0^1 [f(x) - \phi(x)]dx$$

which are integrated separately, the first part by mathematical theory and the second part by crude Monte Carlo. This amounts to estimating a correction [f(x)-ϕ(x)] to the function $\phi(x)$, which approximates f(x).

Denoting: $\int_0^1 \phi(x)dx = \Phi$, the used estimator would be:

$$u(x) = \frac{f(x) - \phi(x)}{p(x)} + \Phi \qquad (11)$$

with p(x) as a uniformly distributed probability density function.

The following choices for $\phi(x)$ can be considered:

$$\phi_0(x) = 1$$
$$\phi_1(x) = 1 + x$$
$$\phi_2(x) = 1 + x + \frac{x^2}{2}$$
$$\phi_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$$
$$\phi_4(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$$

which are in fact successive approximations to the integrand e^x. Sampling from these ϕ's can be carried out by use of the rejection method.

The corresponding Φ's are:

$$\Phi_0 = 1$$
$$\Phi_1 = 1.5$$
$$\Phi_2 = \frac{5}{3}$$
$$\Phi_3 = \frac{41}{24}$$
$$\Phi_4 = \frac{206}{120}$$

A procedure for the application of the correlated sample tevhnique is shown in Fig. 6.

```
!          correlated1 f90
!          Correlated Sampling or Control Variates Monte Carlo
!          Variance estimation using recursive formula
!          I=Integral[e**x,0,1]=1.718281828
!          M. Ragheb
           program correlated1
           real mean_value, t, ta
           real :: trials=50
!          Define sampled function
           g(x) = exp(x)
!  Define control variate function
           f1(x) = 1.0 + x
!          Analytically integrated value of f1(x)
           phi1 = 1.5
!          Open output file
           open(44, file = 'random_out')
!          Initialize variables
           t=0.0
           s_square_old = 0.0
!          Sample function
           do k=1,trials
                   xk=k
                   call random(rr)
!          Evaluate control variate
                   s = (g(rr) - f1(rr)) + phi1
                   if (k.GE.2) then
                           ta = t/(xk-1.0)
                           s_square_new =    (xk-2.0)/(xk-1.0)*s_square_old              &
&                                              + (ta-s)*(ta-s)/xk
                           s_square_old = s_square_new
                   end if
                   t=t+s
           end do
!          Estimate mean value, variance, standard deviation,
!          and fractional standard deviation
           mean_value = t/trials
           std_dev = sqrt (s_square_new)
           std_error = std_dev/sqrt(trials)
           frac_stan_dev = std_error/mean_value
!  Write results
           write(*,100) mean_value, std_error, frac_stan_dev, trials
           write(44,100) mean_value, std_error, frac_stan_dev, trials
100        format(1x,'mean value =',e14.8,/,1x,'standard error =',e14.8,/,  &
&              1x,'fractional standard deviation =',e14.8,/,                 &
&                     1x,'number of trials =',e14.8)
           end
```

Figure 6. Procedure for the application of the correlated sampling technique.

A drastic decrease in variance compared with analog and crude sampling was detected by going from a good to a better approximation for f(x), some of the results being shown in the following Table 3.

Table 3. Mean Values and Standard Deviations using Correlated Sampling.

ϕ	$\theta \pm \frac{\sigma}{\sqrt{n}}$	$\theta \pm \sigma$	fsd $\left[\frac{\sigma / \sqrt{n}}{\theta}\right]$	n

$\phi_1(x) = 1 + x$	1.760786 ± 0.006713 1.765383 ± 0.022005	1.760786 ± 0.212283 1.765383 ± 0.220050	0.003813 0.012465	1,000 100
$\phi_2(x) = 1 + x + \frac{x^2}{2}$	1.732800 ± 0.002021 1.737272 ± 0.006795	1.732800 ± 0.063909 1.737272 ± 0.067950	0.001109 0.003911	1,000 100
$\phi_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$	1.722074 ± 0.001728	1.722074 ± 0.015931	0.001000	85
$\phi_4(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$	1.719504 ± 0.000466	1.719504 ± 0.003295	0.000271	50

6. IMPORTANCE SAMPLING

We have:

$$\begin{aligned} \theta &= \int_0^1 f(x)dx \\ &= \int_0^1 \frac{f(x)}{g(x)}\{g(x)\}dx \\ &= \int_0^1 \frac{f(x)}{g(x)}dG(x) \end{aligned} \tag{13}$$

for any functions g and G where:

$$dG(x) = g(x)dx;$$

and if g is positive valued and normalized such that:

$$G(1) = \int_0^1 g(x)dx = 1.$$

If η is a random number sampled from the distribution G, then:

$$\theta = E\left[\frac{f(\eta)}{g(\eta)}\right] \tag{14}$$

Thus, an equivalent estimator would be:

$$v(x) = \Phi.\frac{f(x)}{\phi(x)} \tag{15}$$

with the sampling carried out from the probability density function:

$$p(x) = \frac{\phi(x)}{\Phi}.$$

```
!           importance f90
!           Importance Sampling Monte Carlo
!           Variance estimation using exponential easy function and recursive formula
!           I=Integral[e**x,0,1]=1.718281828
!           M. Ragheb
            program importance
            real mean_value, t, ta
            real :: alpha=0.90
            real :: trials=50
!           Define sampled function
            f(x) = exp(x)
!           Define importance function(normalized)
            g(x) =(alpha/(exp(alpha)-1.0))*exp(alpha*x)
!           Open output file
            open(44, file = 'random_out')
!           Initialize variables
            t=0.0
            s_square_old = 0.0
!           Sample function
            do k=1,trials
                        xk=k
!           Sample Importance Function g(x)
                        call random(rr)
                        x=log(1.0+((exp(alpha)-1.0)*rr))/alpha
!           Evaluate importance estimator
                                    s = f(x)/g(x)
                                    if (k.GE.2) then
                                                ta = t/(xk-1.0)
                                                s_square_new =     (xk-2.0)/(xk-1.0)*s_square_old              &
&                                                                  + (ta-s)*(ta-s)/xk
                                                s_square_old = s_square_new
                                    end if
                        t=t+s
            end do
!           Estimate mean value, variance, standard deviation,
!           and fractional standard deviation
            mean_value = t/trials
            std_dev = sqrt (s_square_new)
            std_error = std_dev/sqrt(trials)
            frac_stan_dev = std_error/mean_value
!   Write results
            write(*,100) mean_value, std_error, frac_stan_dev, trials
            write(44,100) mean_value, std_error, frac_stan_dev, trials
100         format(1x,'mean value =',e14.8,/,1x,'standard error =',e14.8,/,  &
&                1x,'fractional standard deviation =',e14.8,/,                  &
&                      1x,'number of trials =',e14.8)
            end
```

Figure 7. Procedure for the application of the Importance Sampling technique using the exponential function.

Results of computations are shown in the following Table 4 for the cases:

$$\phi_1(x) = 1 + x$$

$$\phi_2(x) = 1 + x + \frac{x^2}{2}$$

$$\phi_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$$

$$\phi_4(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$$

as successive approximations of the integral. The rejection method was used to sample these functions. Better results than for correlated sampling were obtained.

Table 4. Mean Values and Standard Deviations for Importance Sampling.

Φ	$\theta \pm \frac{\sigma}{\sqrt{n}}$	$\theta \pm \sigma$	*fsd* $\left[\frac{\sigma/\sqrt{n}}{\theta}\right]$	n
$\phi_1(x) = 1 + x$	1.722270 ± 0.005085 1.725310 ± 0.016505	1.722270 ± 0.160801 1.725310 ± 0.165050	0.002952 0.009566	1,000 100
$\phi_2(x) = 1 + x + \frac{x^2}{2}$	1.719404 ± 0.001718 1.721177 ± 0.004604	1.719404 ± 0.044436 1.721177 ± 0.046040	0.001000 0.002675	669 100
$\phi_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$	1.720331 ± 0.001584	1.720331 ± 0.011200	0.001000	50
$\phi_4(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$	1.718664 ± 0.000300	1.718664 ± 0.002121	0.000174	50

7. WEIGHTED UNIFORM SAMPLING

We use the estimator:

$$w_n(\xi_1, \xi_2, \ldots, \xi_3) = \frac{\Phi . \sum_{i=1}^{n} g(\xi_i)}{\sum_{i=1}^{n} \gamma(\xi_i)} = \frac{N_n}{D_n} \tag{16}$$

where the ξ_i's are independent random numbers with identical probability density function p(s):

$$g(s) = \frac{f(s)}{p(s)} \quad \text{where } p(s) \# 0$$

$$\infty \quad \text{where } p(s) = 0$$

and:

$$\gamma(s) = \frac{\phi(s)}{p(s)} \quad \text{where } p(s) \# 0$$

$$\infty \quad \text{where } p(s) = 0$$

The following approximating functions were used:

$$\gamma_1(x) = 1 + x$$

$$\gamma_2(x) = 1 + x + \frac{x^2}{2}$$

$$\gamma_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$$

$$\gamma_4(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$$

Results of computations are shown in Table 5 below.

Table 5. Mean Values and Standard Deviations for Weighted Uniform Sampling.

γ	$\theta \pm \frac{\sigma}{\sqrt{n}}$	$\theta \pm \sigma$	*fsd* $\left[\frac{\sigma/\sqrt{n}}{\theta}\right]$	n
$\gamma_1(x) = 1 + x$	1.755934 ± 0.000489 1.786638 ± 0.003492	1.755934 ± 0.015463 1.786638 ± 0.034920	0.000278 0.001954	1,000 100
$\gamma_2(x) = 1 + x + \frac{x^2}{2}$	1.730234 ± 0.000172 1.739738 ± 0.001060	1.730234 ± 0.005061 1.739738 ± 0.010600	0.000100 0.000609	866 100
$\gamma_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$	1.721938 ± 0.000171 1.723457 ± 0.000090	1.721938 ± 0.002460 1.723457 ± 0.000900	0.000100 0.000165	207 100
$\gamma_4(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$	1.719597 ± 0.000090	1.719597 ± 0.000636	0.000052	50

8. ANTITHETIC VARIATES

In the Antithetic Variates method, variance reduction is achieved by symmetrization of the integral through a group transformation.

An estimator that can be used is:

$$t_1 = \frac{1}{2}\left[f(\xi) + f(1-\xi) \right] \qquad (17)$$

which amounts to an average of the function f(ξ) and its mirror image f(1-ξ). As shown in Fig. 2, less variation from the mean can be expected from the estimator t_1 than by using f(ξ) or f(1-ξ) alone.

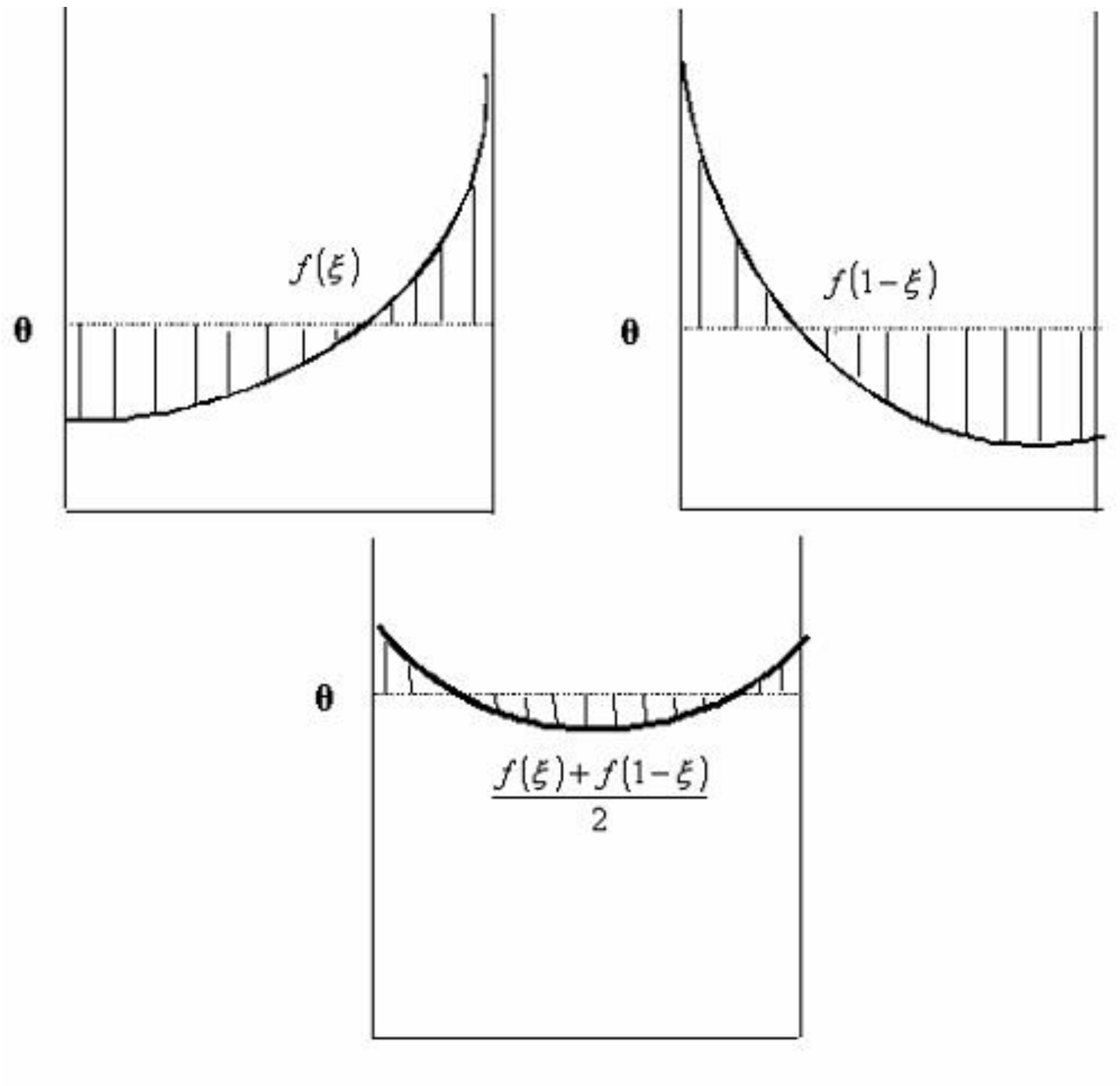

Figure 8. Variance reduction through the use of the Antithetic Variates estimator:

$$t_1 = \frac{1}{2}\left[f(\xi) + f(1-\xi) \right]$$

9. STRATIFICATION

Another estimator based on the Antithetic Variates principle uses stratification of the integration interval such as:

$$t_2 = \frac{1}{4}\left[f\left(\frac{1}{2}\xi\right) + f\left(\frac{1}{2} - \frac{1}{2}\xi\right) + f\left(\frac{1}{2} + \frac{1}{2}\xi\right) + f\left(1 - \frac{1}{2}\xi\right) \right] \qquad (18)$$

```
!     antithetic2 f90
!     Antithetic Variate Monte Carlo
!     Variance estimation using recursive formula
!     I=Integral[e**x,0,1]=1.718281828
!     M. Ragheb
      program importance
      real mean_value, t, ta
```

```
        real :: trials=50
!       Define sampled function
        f(x) = exp(x)
!       Open output file
        open(44, file = 'random_out')
!       Initialize variables
        t=0.0
        s_square_old = 0.0
!       Sample function
        do k=1,trials
                xk=k
!       Sample Importance Function g(x) using rejection
                call random(rr)
!       Evaluate antithetic variates estimator
!               Antithetic variate
!                       s = (f(rr)+f(1.0-rr))/2.0
!               Stratification
                        s=(f(rr/2.0)+f(0.5-rr/2.0)+f(0.5+rr/2.0)+f(1.0-rr/2.0))/4.0
                        if (k.GE.2) then
                                ta = t/(xk-1.0)
                                s_square_new =  (xk-2.0)/(xk-1.0)*s_square_old          &
&                                               + (ta-s)*(ta-s)/xk
                                s_square_old = s_square_new
                        end if
                t=t+s
        end do
!       Estimate mean value, variance, standard deviation,
!       and fractional standard deviation
        mean_value = t/trials
        std_dev = sqrt (s_square_new)
        std_error = std_dev/sqrt(trials)
        frac_stan_dev = std_error/mean_value
!   Write results
        write(*,100) mean_value, std_error, frac_stan_dev, trials
        write(44,100) mean_value, std_error, frac_stan_dev, trials
100     format(1x,'mean value =',e14.8,/,1x,'standard error =',e14.8,/,  &
&           1x,'fractional standard deviation =',e14.8,/,               &
&                 1x,'number of trials =',e14.8)
        end
```

Figure 9. Procedure for the application of the Antithetic Variates and Stratification Sampling technique with the t_1 and t_2 estimators.

Results are shown in Tables 6 and 7, and demonstrate the considerable advantage obtained over other variance reduction methods.

Table 6. Mean Values and Standard Deviations for Antithetic Variates.

Estimator t	$\theta \pm \frac{\sigma}{\sqrt{n}}$	$\theta \pm \sigma$	*fsd* $\left[\frac{\sigma/\sqrt{n}}{\theta}\right]$	n
$t_1 = \frac{1}{2} f(\xi) + \frac{1}{2} f(1-\xi)$	1.717953 ± 0.000440	1.717953 ± 0.062225	0.000256	20,000
	1.715290 ± 0.001964	1.715290 ± 0.062107	0.001145	1,000
	1.717494 ± 0.006554	1.717494 ± 0.065540	0.003816	100

Table 7. Mean Values and Standard Deviations for Stratification.

Estimator t	$\theta \pm \frac{\sigma}{\sqrt{n}}$	$\theta \pm \sigma$	*fsd* $\left[\frac{\sigma/\sqrt{n}}{\theta}\right]$	n
$t_2 = \frac{1}{4} f\left(\frac{1}{2}\xi\right) + \frac{1}{4} f\left(\frac{1}{2} - \frac{1}{2}\xi\right) + \frac{1}{4} f\left(\frac{1}{2} + \frac{1}{2}\xi\right) + \frac{1}{4} f\left(1 - \frac{1}{2}\xi\right)$	1.717516 ± 0.000499 1.718074 ± 0.001667	1.717516 ± 0.015779 1.718074 ± 0.016670	0.000291 0.000970	1,000 100

10. VARIANCE, LABOR AND EFFICIENCY RATIOS

The standard error in Monte Carlo simulations is expressed as:

$$\varepsilon \ \ \alpha_\ell \ \frac{\sigma}{\sqrt{N}} \tag{19}$$

where σ is the variance of the mean and N is the number of trials.

The objective of a Monte Carlo simulation is to obtain an unbiased result with a small standard error in the final estimate of the mean. It is customary to reduce the error to increase the number of trials N. It can be seen from Eqn. 19 that to reduce the error by a factor of k, the number of trials must be increased by a factor of k^2. For instance to decrease the error by a factor of 2, the number of trials must be increased by a factor of 4. This becomes very impracticable when the value of the needed reduction in the error k is large, making many problems practically unsolvable of the most powerful of computer platforms.

The appropriate approach is to reduce the error by a proper choice of the methods in which the numerical experiment is planned and carried out, which amounts to a reduction of the factor σ in the denominator of Eqn. 19. This could entail extra computations and the use of different estimators.

To compare two Monte Carlo sampling schemes 1 and 2 with different variances, we define the variance ratio:

$$V_{12} = \frac{\sigma^2{}_2}{\sigma^2{}_1} \tag{20}$$

To compare two sampling schemes with different labor expressed in terms of the computational time T, we define the labor ratio:

$$L_{12} = \frac{T_2}{T_1} \tag{21}$$

To account for *both* the extra labor and computational time T, we define the efficiency of a Monte Carlo sampling scheme 1 as:

$$\eta_1 \ \ \alpha_\ell \ \frac{1}{\sigma^2{}_1 T_1} \tag{22}$$

For another sampling scheme 2, the efficiency is:

$$\eta_2 \ \alpha_\ell \ \frac{1}{\sigma^2{}_2 T_2} \tag{23}$$

These efficiencies are considered to be inversely proportional to the product of the sample variance and the amount of labor expended in the computation expressed in terms of the computational time T. This extra labor is normally caused by the use of a more sophisticated sampling scheme.

To compare two sampling methods 1 and 2, one can define a figure of merit F or efficiency ratio in terms of the ratio of their respective efficiencies as:

$$F_{12} = \frac{\eta_1}{\eta_2} = V_{12} L_{12} = \frac{\sigma^2{}_2 T_2}{\sigma^2{}_1 T_1} \tag{24}$$

It can be noticed that the efficiency ratio is the product of the variance and labor ratios. The comparison in terms of the efficiency ratio F_{12} can be extended to the comparison of the Monte Carlo method to another numerical method if its error estimate can be obtained.

An alternate definition of the efficiency ratio can be written in terms of the standard deviations rather than the variances as:

$$F'_{12} = \frac{\eta_1}{\eta_2} = S_{12} L_{12} = \frac{\sigma_2 T_2}{\sigma_1 T_1} \tag{25}$$

with a standard deviation ratio defined as:

$$S_{12} = \frac{\sigma_2}{\sigma_1} \tag{26}$$

11. OVER-BIASING AND UNDER-BIASING: CHARACTERISTIC MONTE CARLO PITFALLS

In modifying the sampling for variance reduction, i.e. improved efficiency or making the best use of computing time to stimulate events which are most significant to the final answer, it is possible to overshoot the mark and produce a sampling scheme that is so strongly biased as to be less efficient than crude sampling. This is over-biasing or over-sampling.

The opposite situation is also encountered as under-sampling or under-biasing. This occurs in crude or a slightly modified sampling scheme when the result depends heavily on infrequent events and not enough observations occurred for good statistics.

It is a general characteristic of both cases that most of the time the generated answers are too small. This produces an apparently consistent bias in the result. Variance estimates are also generally small so that the confidence intervals calculated in the simulation will tend to indicate that the results are much more accurate than they really are. This generates a false feeling of security and faith in the obtained results, which are actually consistently bad.

McGrath and Irving demonstrate these ideas by an extremely simplified example: Consider a simulation with two types of events. One type (X_1) occurs frequently ($f(X_1) = 0.999$) but contributes only a small amount ($g(X_1) = 0.01$) to the final result, while the other type (X_2) is rare ($f(X_2) = 0.0001$) but makes a large contribution ($g(X_2) = 100$) when it occurs.

The quantity I being estimated has the correct value:

$$\begin{aligned} I &= f(X_1)g(X_1) + f(X_2)g(X_2) \\ &= 0.9999x0.01 + 0.0001x100 \\ &= 0.02 \end{aligned} \tag{27}$$

Using Crude Monte Carlo with a moderate number of histories, event X_2 would rarely occur and an *underbiased* answer would be recorded as:

$$I_u = I\ g(X_1) = 0.01 \tag{28}$$

If importance sampling is used so that the X_2 events occurred frequently; since they contribute so much to the answer, (to an excess), say new probabilities of $f^*(X_2) = 0.9999$ and $f^*(X_1) = 0.0001$, were used, then the X_1 events will never occur in a run of moderate size and the *overbiased* result will be:

$$I_0 = g\left(X_2\right)\frac{f(X_2)}{f^*(X_2)} = 100.0\frac{0.0001}{0.9999} = 0.01 \tag{29}$$

The proper modification for that case is to let X_1 and X_2 happen with equal probabilities, $f^*(X_1) = f^*(X_2) = 0.5$ (according to their contribution to the final answer), then the contribution from each history is:

$$\begin{aligned} & g(X_1)\frac{f(X_1)}{f^*(X_1)} = 0.01\frac{0.9999}{0.5} = 0.02 \\ & = g(X_2)\frac{f(X_2)}{f^*(X_2)} = 100\frac{0.0001}{0.5} = 0.02 \end{aligned} \tag{30}$$

and the final estimate from a small sample would be:

$$I = 0.02$$

The underlying distribution is highly skewed with the large majority of cases making little or no contributions to the final answer, while a small number of cases can make large contributions.

In both cases the final estimates were smaller than the correct answer, which is a general characteristic of such cases. If a set of histories, say 100, was simulated by crude Monte Carlo, then most likely there would be no X_2 events observed and the (incorrect) estimate will be 0.01. Once every 100 sets of 100 histories, a single event X_2 will happen and the estimate will be:

$$I_u^{'} = \frac{99 \times 0.01 + 1 \times 100}{100} = 1.01, \qquad (31)$$

a number very much larger than the correct value. Even though the estimation procedure is right and averages come out correctly in the long run, most users would think this was the result of some input mistake or computation error, and throw out the run. Therefore caution is recommended in simulations where most histories contribute a small bit to the answer but a few histories contribute a large value. Complete faith should not be placed in estimates of variance especially when the results are smaller than expected or if the possibility of overbiasing is suspected.

EXERCISES

1. Using the Crude Monte Carlo method, write a procedure to estimate the value of the integral: $\int_0^1 (1+x)dx$. Use an equation for the estimation of the variance of your choice. Plot the mean value and the variance as a function of the number of experiments N.
2. Using the correlated sampling variance reduction method, write a procedure to estimate the value of the integral $\int_0^1 (1+x)dx$ using $\int_0^1 e^x dx$ as an easy function. Plot the mean value and the variance as a function of the number of experiments N. Compare the result to that obtained from the crude sampling Monte Carlo method.
3. Compare the results for the Correlated Sampling method using the different approximation functions discussed.
4. Derive a higher level Antithetic Variates function t_3, and compare its results to those obtained using the functions t_1 and t_2.
5. Write a procedure using weighted Uniform Sampling for estimating an integral. Discuss any produced bias in the result.
6. Compare the mean values, standard errors, fractional standard deviations, variance ratios, and efficiency ratios for the Crude Monte Carlo Sampling method for the estimation of an integral of your choice for different numbers of trials. For instance you could use the integral of $\int_0^1 e^x dx$ over the interval [0,1]. Now use the Antithetic Variates estimators t_1 and t_2, and compare their results to each other and to those obtained using the Crude Monte Carlo estimator. Choose the variance estimation formula of your choice. Use the difference between two calls to the computer clock at the beginning and end of a calculation to estimate the labor time.