POLYTECHNIC UNIVERSITY
Department of Computer and Information Science

VARIANCE REDUCTION TECHNIQUES IN MONTE CARLO SIMULATIONS

K. Ming Leung

Abstract: Techniques for reducing the variance in Monte Carlo simulations are discussed. The basic ideas will be illustrated using the rejection method and the importance sampling method.

Directory

Last Revision Date: March 9, 2004 Version 2.0

Table of Contents

1. Introduction

Monte Carlo simulation provides an estimation of the mean value by randomly sampling the population. If the number of trials is N, then the error is always characterized by an $N^{-1/2}$ dependence. This is a rather slow convergence, because if we want to further decrease the error say by a factor of 10, then we have to increase the number of trials by a factor of 100.

As we noted before, the error is also proportional to the square root of the variance of the function we are integrating, and the variance has to do with the way how sampling is carried out. So far we have only considered random and unbiased sampling of the population. In this chapter, we will discuss methods in which we cheat a little and sample the population in some carefully chosen biased fashion so as to reduce the variance of the method and therefore to lower the simulation error. These methods are referred to as variance reduction methods. This biased sampling will be done more effectively over the most important part of the population (the integrand) by appropriate use of non-uniform random deviates. In the case of evaluating an

integral of $\rho(x)$ from a to b

$$I = \int_a^b \rho(x)\, dx, \tag{1}$$

by simulation, the most important part corresponds to the part of the interval $[a, b]$ where $\rho(x)$ is large and so contributes most to the integral. Because of this, the method is also known as importance sampling.

2. Motivations for Variance Reduction

As we saw before the variance depends very much on the behavior of the function, $\rho(x)$ being integrated. If $\rho(x)$ varies little in magnitude within the domain of integration, then the variance is small. The estimated mean from a uniform sampling of the integrand gives a very accurate result for the integral. In the limit that $\rho(x)$ is a constant, the variance goes to 0, the estimated mean gives the exact answer for I. On the other hand if $\rho(x)$ varies significantly over the domain then the variance is large, and so is the expected error in estimating the

mean.

It is easy to see why the error behave as so. For example in the sample mean Monte Carlo method, the function is sampled uniformly within the interval $[a, b]$. Each sample point, x_n contributes a term $\rho(x_n)/N$ to the estimate value of the mean, which when multiplied by $b - a$ and summed gives an estimation of the integral:

$$I' = (b-a)\mu_{\bar{Y}} = \frac{b-a}{N}\sum_{n=1}^{N}\rho(x_n). \tag{2}$$

If $\rho(x)$ varies in magnitude a lot within $[a, b]$, then the sample points falling near the minimum of the function contribute much less to the estimated mean than points that fall near the peak of the function. Sample points like these are to some extend wasted in the simulation and the number of them should be minimized as much as possible. It is clear that we should bias the way we sample the function so that more points will be sampled near the peak of the function. This type of sampling strategy is referred to as importance sampling.

3. Variance Reduction using the Rejection Method

One way to sample the function more near it peak than in the region where it is small is to use the rejection method. The fact that $\rho(x)$ is in general not normalized does not affect our argument here.

As we discussed before in the reject method, in the graph of $\rho(x)$ versus x, if points can be generated so as to cover the area under the curve within the domain, then the value of x of these points will be distributed according to $\rho(x)$. The question is how can these points be generated efficiently? In the rejection method, we choose a comparison function $w(x)$, which has a finite area, A, between a and b, and within that domain $w(x) \geq \rho(x)$. We also require that the indefinite integral

$$W(x) = \int_{-\infty}^{x} w(x)\, dx \tag{3}$$

can be obtained analytically, and $Au = W(x)$ can be analytically inverted to give $x = W^{-1}(u)$. Here u is a uniform deviate. That means that the inverse function method can be used to produce random deviates that are distributed according to $w(x)$. Thus points can be

generated uniformly covering the area beneath the $w(x)$ curve. Out of all these points, the ones lying above the $\rho(x)$ curve are then rejected. The remaining points must therefore lie underneath the curve.

If the total number of trials is N, out of which N' points are accepted, then we can approximate the integral as

$$I \approx A\frac{N'}{N}. \tag{4}$$

The reasoning here is exactly the same as that used for the Hit-Or-Miss method.

Specifically, for each point x_n lying beneath $w(x)$, we let $y_n = w(x_n)u_n$, where u_n are uniform deviates. We are interested in a discrete random variable Q, whose values q are given by

$$q_n = \begin{cases} 1, & \text{if } y_n \leq \rho(x_n) \\ 0, & \text{if } y_n > \rho(x_n) \end{cases} \tag{5}$$

As in the Hit-Or-Miss method, if we denote the ratio of the area under $\rho(x)$ to the area under $w(x)$ by

$$r = \frac{I}{A}, \tag{6}$$

then the discrete variable Q obeys the following probability distribution:

q	1	0
$p(q)$	r	(1-r)

As we have seen before, the mean $E(Q) = r$ and the variance $V(Q) = r(1-r)$. The simulation gives $I = Ar = AE(Q)$ and the probable error for I is given by

$$I_{\text{error}}^{\text{RJ}} = 0.67\sqrt{\frac{I(A-I)}{N}}. \tag{7}$$

This equation looks identical to the one we obtained before for the Hit-Or-Miss method, except the area A is the area under the $w(x)$ curve and does not have to be the area of an inscribing rectangle.

3.1. Algorithm for the Generalized Rejection Method

The procedure for using the generalized rejection method to compute an integral is summarized below.

1. Set N to a positive integer, and N' to 0.
2. Go through the following loop N times:
   (a) Let $x = W^{-1}(Au_1)$, where u_1 is a uniform deviate.
   (b) Get another uniform deviate u_2 and let $y = u_2 w(x)$.
   (c) If $y < f(x)$, retain the value of x and increment N' by 1, otherwise reject the point (x, y).
3. Approximate I by $I' = AN'/N$.
4. Probable error is estimated as $0.67\sqrt{I'(A - I')/N}$.

3.2. Application of the Rejection method

We will apply the above technique to compute the integral of the sine function

$$I = \int_0^{\pi/2} \sin x \; dx. \tag{8}$$

As x goes from 0 to $\pi/2$, the sine function goes from 0 to to a peak value of 1. We will use the comparison $w(x) = x$ because $x \geq \sin x$ for all x in $[0, \pi/2]$, and the inverse function method can be used to generate random deviates distributed according to this $w(x)$. Notice that we cannot directly use the results we derived previously for a probability density function varying as a power of x. Because we assumed the domain to be the interval $[0, 1]$. Of course we can scale the results to fit the present interval of $[0, \pi/2]$. However the calculation is so simple that it is easier to repeat it again here.

We first calculate

$$W(x) = \int_0^x x \; dx = \frac{x^2}{2}. \tag{9}$$

The area under the $w(x)$ curve is given by

$$A = W(\pi/2) = \frac{\pi^2}{8}. \tag{10}$$

We choose a uniform deviate u and set $Au = W(x)$, that is

$$\frac{\pi^2}{8}u = \frac{x^2}{2}. \tag{11}$$

Solving for x in terms of u gives us a formula for generating the non-uniform deviates distributed according to this $w(x)$:

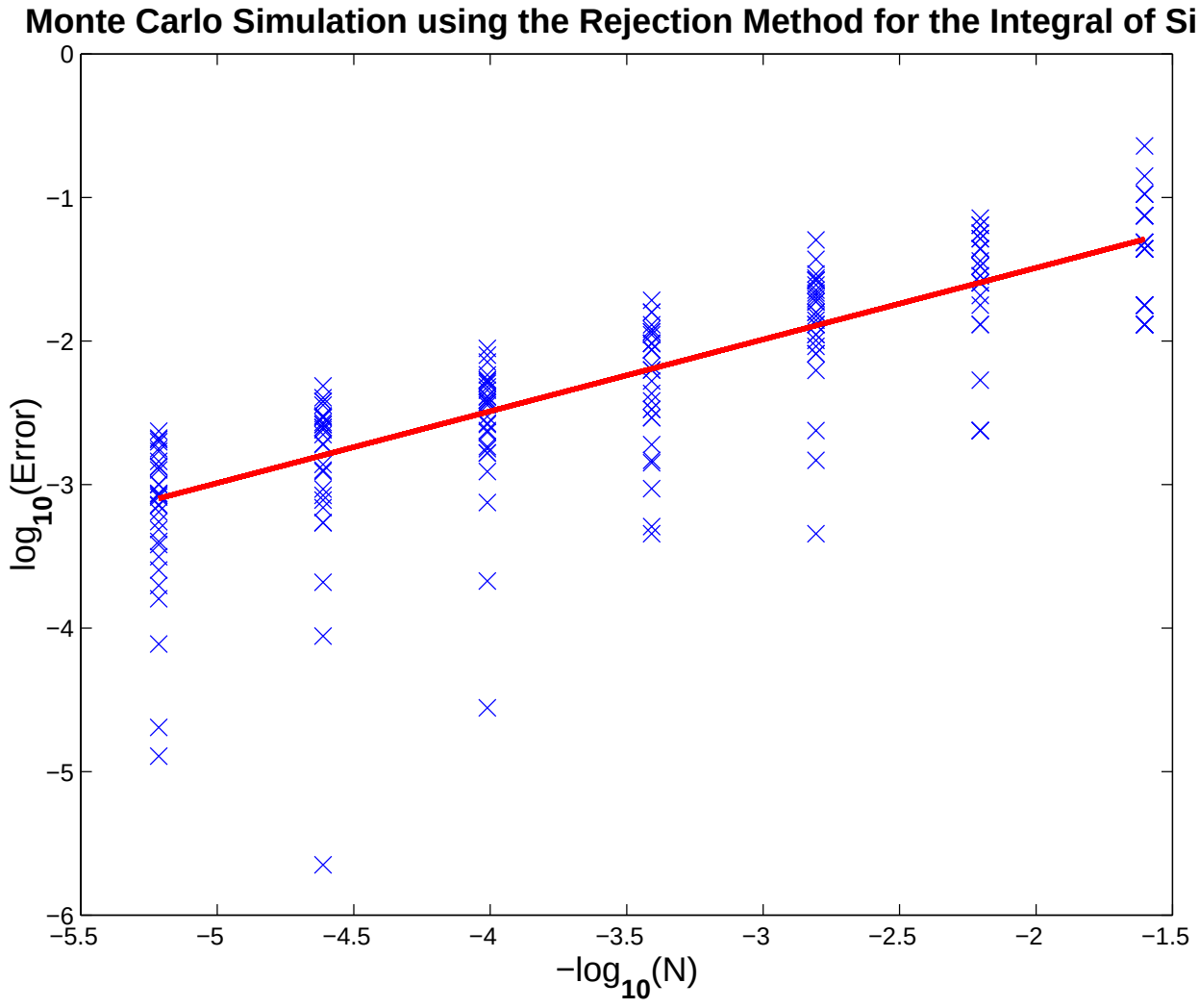
$$x_n = \frac{\pi}{2}\sqrt{u_n}. \tag{12}$$

We can then follow the procedure for the rejection method to compute the integral, I. The estimated value for I can be compared with the exact answer of 1. We can also use the data to estimate the probable error using Eq.(7). According to theory, since $I = 1$ and from Eq.(10) $A = \pi^2/8$, the probable error for the present problem should

be given by

$$I_{\text{error}}^{\text{RJ}} = 0.67\sqrt{\frac{I(\frac{\pi^2}{8} - I)}{N}}. \tag{13}$$

This is exactly the same as the probable error if it is simulated using the sample mean method. For the present problem and the present choice of $w(x)$, we expect the probable error for the rejection method to be the same as that for the sample mean. This agreement is purely accidental. If a slightly different function $w(x)$ is used to cover up $\rho(x)$, the area under $w(x)$, A, will be different, and the agreement will not hold.


Monte Carlo Simulation using the Rejection Method for the Integral of Sine
0
−1
−2
−3
−4
−5
−6
−5.5
−5
−4.5
−4
−3.5
−3
−2.5
−2
−1.5
log10(Error)
−log10(N)

4. Motivation of the Importance Sampling Method

In the sample mean method, the region within the domain of the integral is sampled in a uniform and unbiased way. If the integrand varies in magnitude significantly within the domain then sample points that fall near the minimum of the function contribute little to the integral and to some extend are wasted in the simulation.

On the other hand, if the integrand does not vary much in magnitude within the domain, then each sample point contributes almost equally to the integral. In the limit that the integrand is a constant, then one can obtain the exact value of the mean even with a single sample point, x_1. The resulting estimation of the integral therefore has no error:

$$I' = \frac{b-a}{N} \sum_{n=1}^{N} \rho(x_n) = (b-a)\rho(x_1), \tag{14}$$

because the area of interest is then a rectangle whose height is given exactly by $\rho(x_1)$.

If one looks at the integral for I, the integrand is given by $\rho(x)$.

How can we change it so that the new integrand becomes smoother? There is no other alternatives except to try to make an appropriate transformation of the variable of integration.

5. Important Sampling

We now want to consider transforming the variable of integration in such a way that more points will sample the more significant part of the integrand. This method of variance reduction is referred to as importance sampling.

To see how to make the necessary transformation, we start with a function $f(x)$ that is non-negative within the interval $[a, b]$, and its indefinite integral, which we denote by $F(x)$, can be computed

$$F(x) = \int_{-\infty}^{x} f(x)\, dx. \tag{15}$$

Note that $f(x)$ does not have to be normalized. If it is normalized then it can be interpreted as a probability density function.

Next let us transform the integration variable from x to y via the

relation

$$y = F(x). \tag{16}$$

Differentiating this equation we have $\frac{dy}{dx} = f(x)$ and so $dx = \frac{dy}{f(x)}$. We can invert the above relation to express x in terms of y: $x = F^{-1}(y)$. The integral, I, can now be rewritten as

$$I = \int_a^b \rho(x)\, dx = \int_a^b \frac{\rho(x)}{f(x)}\, dy = \int_{F(a)}^{F(b)} \frac{\rho(F^{-1}(y))}{f(F^{-1}(y))}\, dy. \tag{17}$$

Now specifically what should we pick for $f(x)$? It is clear that if we choose it to be a constant c times $\rho(x)$, then in that case the integrand, when expressed in terms of the variable y, is exactly given by a constant $1/c$, and we have seemingly achieved what we intended to achieve. However, a closely look reveals that the solution for I in Eq. (17) is given by

$$I = \frac{1}{c}(F(b) - F(a)), \tag{18}$$

but the function $F(x)$ is defined in terms of an integral of $f(x)$, which in this case involves an integral of $\rho(x)$. But this is what we want to

find in the first place. More precisely we find,

$$I = \frac{1}{c}(F(b) - F(a)) = \frac{1}{c}\left[\int_{-\infty}^{b} - \int_{-\infty}^{a}\right] f(x)\, dx = \int_{a}^{b} \rho(x)\, dx. \quad (19)$$

Thus we are back to where we started.

Of course we do not have to be so ambitious as to choose a function $f(x)$ so as to make the integrand exactly constant, as we attempted above. The strategy for reducing the variance of our simulation is now very clear. What we need is to pick a non-negative function $f(x)$ which is large where $\rho(x)$ is large and is small where $\rho(x)$ is small. In that case the integrand is then rather smooth after the above transformation is made. Moreover the function $f(x)$ should be such that it's indefinite integral can be calculated to obtain $F(x)$, and in addition F^{-1} can be found by inverting F analytically, or accurate and efficient numerical approximations for F^{-1} is available.

Instead of looking at it as a transformation of the integration variable, there is an equivalent view of Eq. (17). First note that the overall scale for $f(x)$ is irrelevant. For example, changing it by a scale

factor of α means that dy is changed by exactly the same factor and so the equation does not change. Therefore we can assume that $f(x)$ is normalized, and so it can be considered as a probability density function. Next, since $dy = f(x)\ dx$ we can write Eq. (17) as

$$I = \int_a^b \frac{\rho(x)}{f(x)}\ dy = \int_a^b \left[\frac{\rho(x)}{f(x)}\right] f(x)\ dx. \tag{20}$$

We also define an average or expected value of an arbitrary function $\eta(x)$ by

$$< \eta >_f = \int_a^b \eta(x) f(x)\ dx. \tag{21}$$

It is the average of the $\eta(x)$ weighted by a probability density function $f(x)$. Therefore in terms of the function

$$\gamma(x) = \frac{\rho(x)}{f(x)}, \tag{22}$$

we have

$$I = \int_a^b \gamma(x) f(x)\ dx = < \gamma >_f\ . \tag{23}$$

Again the idea is to choose a probability density function $f(x)$ that is large where $\rho(x)$ is large and small where $\rho(x)$ is small, so that $\gamma(x)$ varies little within the domain $[a, b]$. If x_n are random deviates distributed according to the probability density function $f(x)$, then we can estimate the value of $< \gamma >_f$ by the sample average so that the integral can be approximated by

$$I =< \gamma >_f \approx \frac{1}{N} \sum_{n=1}^{N} \gamma(x_n), \tag{24}$$

where N is the number of sample points. The probable error is clearly given by

$$I_{error} = 0.67 \sqrt{\frac{V_f(\gamma)}{N}}, \tag{25}$$

where $V_f(\gamma)$ is the variance of $\gamma(x)$ defined by

$$V_f(\gamma) =< \gamma^2 >_f - (< \gamma >_f)^2. \tag{26}$$

The variance can be estimated from the sampled values $\gamma(x_n)$ because

we can estimate $< \gamma^2 >_f$ by

$$< \gamma^2 >_f \approx \frac{1}{N} \sum_{n=1}^{N} \gamma^2(x_n). \tag{27}$$

Therefore we can estimate the probable error from the sample data.

Theoretically since $< \gamma >_f = I$, the theoretical value for the probable error is given by

$$I^{\text{IS}}_{\text{error}} = 0.67 \sqrt{\frac{< \gamma^2 >_f - I^2)}{N}}. \tag{28}$$

5.1. Importance Sampling Algorithm

The algorithm for MC simulation with importance sampling is:

1. Initialize N to a large integer and sums s_1 and s_2 to 0.
2. Go through the following loop N times:
   (a) Generate points x_n distributed according to $f(x)$.
   (b) Let $\gamma_n = \rho(x_n)/f(x_n)$.
   (c) Add γ_n to s_1.
   (d) Add γ_n^2 to s_2.
3. Estimate the integral $I' = \frac{s_1}{N}$.
4. Estimate the variance $V' = \frac{s_2}{N} - I'^2$.
5. Estimated the probable error for I: $0.67\sqrt{\frac{V'}{N}}$.

5.2. The Special Case of Uniform $f(x)$

If we choose a probability density function $f(x)$ as indicated above, we expect the variance will be reduced. Of course the above formalism

still works even if we choose an $f(x)$ that is small where $\rho(x)$ is large and is large where $\rho(x)$ is small. In that case the variance is expected to be larger than in the sample mean method where the sampling is done uniformly. For the special case where $f(x)$ is a constant, we expect the method to be exactly the same as the sample mean. This is indeed the situation as can be easily checked analytically.

First, for a constant $f(x)$, it must be given for x within the interval $[a, b]$ by $1/L$, where L is the length of the interval. Second for a uniform distribution, the sample points are given by $x_n = Lu_n + a$, where the u_n are uniform deviates. The sampled values are therefore given by $\gamma(x_n) = L\rho(x_n)$, and so the integral is approximated by

$$I \approx \frac{L}{N} \sum_{n=1}^{N} \rho(x_n), \tag{29}$$

which is precisely the same expression as for the sample mean method.

5.3. An Example of Using Importance Sampling

As an example illustrating the use of the importance sampling method to reduce the variance in a Monte Carlo simulation, we will consider again the problem of computing the integral

$$I = \int_0^{\pi/2} \sin x \; dx. \tag{30}$$

For this problem we have $a = 0$, $b = \pi/2$, $L = \pi/2$ and $\rho(x) = \sin x$. Notice that the integrand $\rho(x)$ is small when x is small and has a peak of magnitude 1 at the right end point of $\pi/2$, as shown by the blue solid curve in the figure. The variance of ρ is not small, and we expect that the importance sampling method can significantly improve the accurate of the simulation.

We seek a probability density function $f(x)$ that is small near the left end point and large near the right end point. Moreover $f(x)$ must be simple enough that the inverse function method can be applied to generate random deviates that are distributed according to $f(x)$. We will use a linear function for $f(x)$ and write $f(x) = \alpha x$. The value of

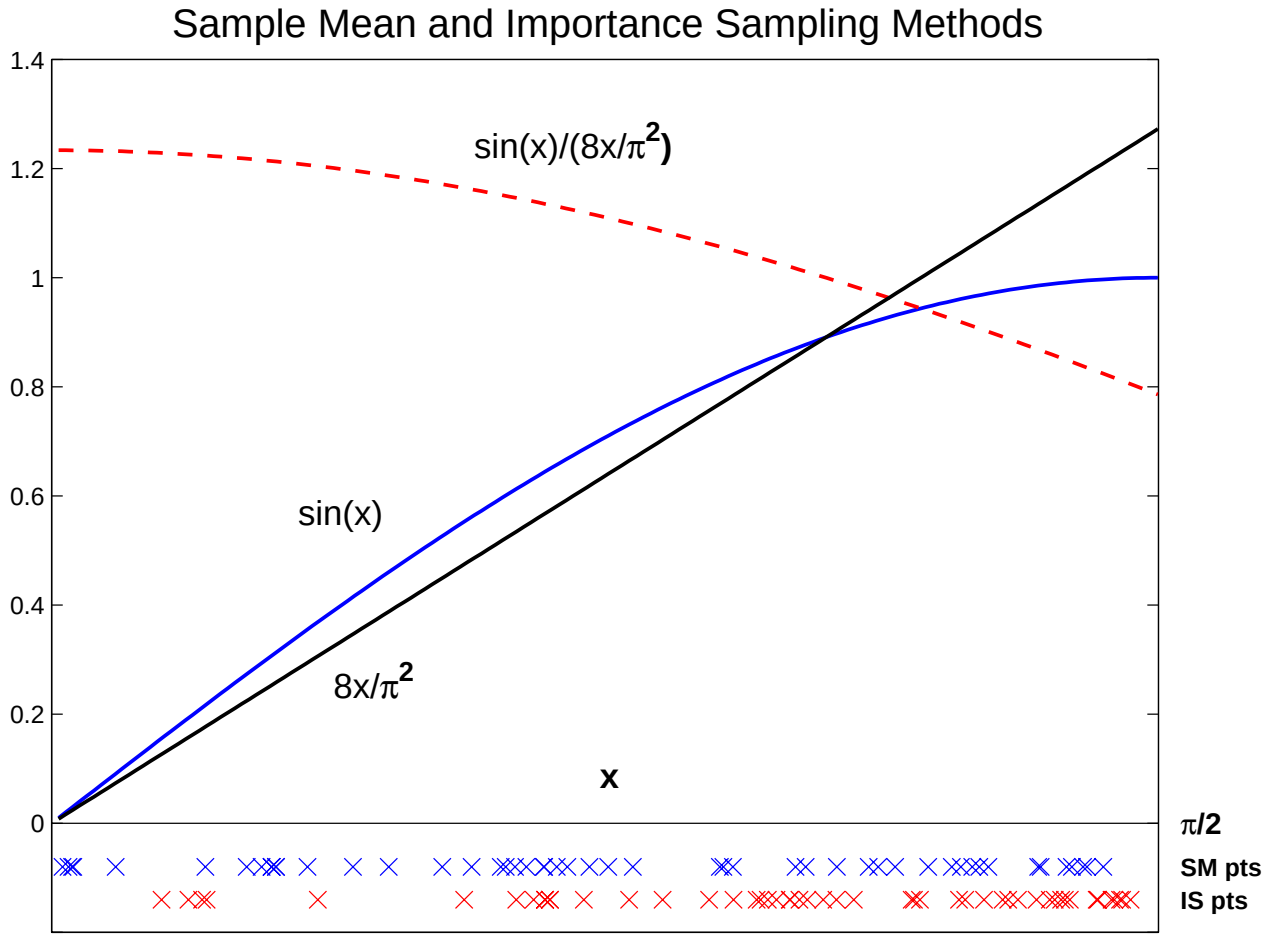
α is determined from normalization of $f(x)$:

$$\int_a^b f(x)\ dx = \alpha \int_0^{\pi/2} x\ dx = \frac{\alpha\pi^2}{8} = 1, \tag{31}$$

and so we have $\alpha = 8/\pi^2$. Therefore $f(x) = \frac{8x}{\pi^2}$ and so

$$\gamma(x) = \frac{\rho(x)}{f(x)} = \frac{\pi^2 \sin x}{8x}. \tag{32}$$

Notice that $\gamma(x)$, as shown by the red dashed curve in the figure, does not vary much within the domain. It takes on its maximum value of about 1.234 at 0 and decreases monotonically to about 0.785 at $\pi/2$.



The cumulative probability distribution function for this density function is given by

$$F(x) = \int_0^x f(x) = \frac{8}{\pi^2} \int_0^x x \, dx = \frac{4x^2}{\pi^2}. \tag{33}$$

Setting this equal to uniform deviates u_n and solving for x gives the non-uniform deviates distributed according to $f(x)$:

$$x_n = \frac{\pi}{2}\sqrt{u_n}. \tag{34}$$

We will use these points to sample the function $\gamma(x)$. The sampled values are then given by $\gamma_n = \frac{\pi^2 \sin x_n}{8x_n}$. Therefore our estimate for the integral is $I' = \frac{1}{N}\sum_{n=1}^{N}\gamma_n$. We can of course also estimate the variance and therefore the probable error from the data.

Before we present the results of our simulation, we want to point out that the present integral has an exact answer of $I =< \gamma >_f = 1$. Moreover we also have

$$< \gamma^2 >_f = \frac{\pi^2}{8}\int_0^{\pi/2}\frac{\sin^2 x}{x}. \tag{35}$$

The above integral is related to the cosine-integral $ci(\pi)$, which has an approximate value of 0.824. Therefore the variance is given by

$$V_f(\gamma) = \approx 0.824\frac{\pi^2}{8} - 1. \tag{36}$$

It is interesting to compare the theoretical values of the probable errors obtained for this integral using the Hit-Or-Miss method, the sample-mean method and the importance sampling method. The ratios of these errors are given by

$$\begin{aligned} I^{\mathrm{HM}}_{\mathrm{error}} : I^{\mathrm{SM}}_{\mathrm{error}} : I^{\mathrm{IS}}_{\mathrm{error}} &= \sqrt{\frac{\pi}{2} - 1} : \sqrt{\frac{\pi^2}{8} - 1} : \sqrt{\frac{0.824\pi^2}{8} - 1} \quad (37) \\ &= 0.756 : 0.483 : 0.129. \end{aligned}$$

Thus the probable error in the important sampling method is about 3.7 times smaller than for the sample mean method. In other words, in order to achieve the same accuracy of the importance sampling method using the sample mean method, about 14 times more sample points must be used. Although the CPU time needed to run a simulation here is proportional to the number of sample points, the importance sampling method is expected to be a little less than 14 times faster than the sample mean method, assuming roughly the same error, since the importance sampling method involves a slightly more complicated integrand.

The figure also shows the points used in the sample mean method. They are distributed rather uniformly in the interval. On the other hand, the points used in the importance sampling method are distributed quite non-uniformly. There are many more points for larger values of x than for smaller values.

6. Comparing the Probable Errors for the Four Methods

Four different methods for computing the integral of a function have been discussed, they are the Hit-Or-Miss method, the sample mean method, the rejection method and the importance sampling method. The probable errors expected of these methods will be analyzed here. We recall the following general expressions for the most probable errors of these methods:

$$I_{\text{error}}^{\text{HM}} = 0.67\sqrt{\frac{I[Lh - I]}{N}}, \tag{38}$$

$$I_{\text{error}}^{\text{SM}} = 0.67\sqrt{\frac{L\int \rho^2\ dx - I^2}{N}}, \tag{39}$$

$$I_{\text{error}}^{\text{RJ}} = 0.67\sqrt{\frac{I(A-I)}{N}}, \tag{40}$$

where A is the area beneath the comparison function.

$$I_{\text{error}}^{\text{IS}} = 0.67\sqrt{\frac{<\gamma^2>_f - I^2}{N}}. \tag{41}$$

Notice that in all of these expressions, the second term in the variance is always given by I^2. For the important sampling method, we note that

$$<\gamma^2>_f = \int_a^b \frac{\rho^2(x)}{f^2(x)} f(x)\ dx = \int_a^b \frac{\rho^2(x)}{f(x)}\ dx. \tag{42}$$

To compare the probable error, we need to compare the following 4 terms:

$$Lh\int_a^b \rho(x)\ dx : L\int_a^b \rho^2(x)\ dx : A\int_a^b \rho(x)\ dx : \int_a^b \frac{\rho^2(x)}{f(x)}\ dx. \tag{43}$$

For the Hit-Or-Miss method, since h cannot be less than the maximum of the function $\rho(x)$ in the domain, its probable error is normally the largest. The rejection method can have a probable error comparable to the sample mean method, depending on how well one can choose the comparison function. The important sampling method is expected to be the best if we have a probability density function $f(x)$ that has a variation in x similar to that of $\rho(x)$.

References

[1] Ilya M. Sobol, *A Primer for the Monte Carlo Method*, (CRC Press, 1994).

[2] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, (Methuen, London, 1964).

[3] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, (Willey, New York, 1986).

[4] P. Bratley, B. L. Fox and E. L. Schrage, *A Guide to Simulation*, (Springer-Verlag, New York, 1983

[5] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery *Numerical Recipes*, Second Edition, Sec. 7.6, p. 295 (Cambridge University Press, 1992).

[6] H. Gould and J. Tobochnik, *An Introduction to Computer Simulation Methods: Applications to Physical Systems*, Second Edition, Ch. 11 (Addison Wesley, New York, 1996).

[7] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery *Numerical Recipes*, Second Edition, Ch. 4, p. 123-155

(Cambridge University Press, 1992).