Topic 1. Simulation and the Monte Carlo Methods

Simulation is a word which is common use today. It has long been an important tool of designers, whether they are simulating a supersonic jet flight, a telephone communication system, a wind tunnel, a large-scale military battle (to evaluate defensive or offensive weapon systems), or a maintenance operation (to determine the optimal size of repair crews).

For statisticians and operation-research workers, the term "simulation" describes a wealth of varied and useful techniques, all connected with the mimicking of the rules of a model of some kind. It frequently occurs that we find processes in the real world far too complicated to understand. The first step in studying complicated processes is to building a model. In such cases it is a good idea to strip the processes of some of their features, to leave us with *models* of the original processes. If we can understand the model, then that may provide us with some insight into the process itself.

The age-structure of the human population of many societies is changing, currently with the fraction of individuals that are elderly increasing. The propect for the future is thus one of progressively more elderly individuals, many of whom will be in need of social services of some kind. Efficient distribution of such services depends upon an understanding of the way in which the population of elderly is composed, and how this structure might change with time.

Quite often the model involves random elements. Because sampling from a particular distribution involves the use of random numbers, stochastic simulation is sometimes call *Monte Carlo simulation*. Historically, the Monte Carlo method was considered to be a technique, using random or pseudorandom numbers, for solution of a model. The term *Monte Carlo* was introduced by von Neumann and Ulam during World War II, as a code word for the secret work at Los Alamos; it was suggused by the gambling casinos at the city of Monte Carlo in Monarco. The Monte Carlo method was then applied to problems related to the atomic bomb. The work involved direct *simulation* of behavior concerned with random neutron diffusion in fissionable material. Shortly thereafter Monte Carlo methods were used to evaluate complex multidimensional integrals and to solve certain integral equations, occurring in physics, that were not amenable to analytic solution.

The Monte Carlo method can be used not only for solution of stochastic problems, but also for solution of deterministic problems. A deterministic problem can be solved by the

Monte Carlo method if it has the same formal expression as some stochastic process.

Chapter 1. Monte Carlo Integration and Variance Reduction Techniques

1 Monte Carlo Integration

Computers can be used to simulate a random sample from a distribution and thereby to estimate parameters of interest defined in terms of that distribution. This approach is referred to as the *Monte Carlo method.* An important advantage of this method over more classical analytic and numerical techniques is that it can readily be applied in complicated situations in which the classical techniques are infeasible.

We now illustrate the use of random samples to estimate various parameters. Let a point (X, Y) be chosen at random from the square $[-1, 1] \times [-, 1]$ and let $R = \sqrt{X^2 + Y^2}$ denote the distance from the chosen point to the center of the square. Suppose we are interested in estimating $E(R)$ and $Var(R)$. We can use the integration technique (which involves the integration of $r^2 arcsecr$) to computer them which turn out to be $E(R) = [\sqrt{2} + \log(1 + \sqrt{2})]/2 \approx 0.7652$ and $Var(R) = [\sqrt{2} + \log(1 + \sqrt{2})]^2/9 \approx 0.08114$.

On the other hand, we can get a random sample of size 100 from the indicated the distribution and calculate their sample mean and sample variance accordingly. If we use S statistical system, it can be done easily as follows:

- Generate Data:

  $n < -100$; $x < -runif(n, -1, 1)$; $y < -runif(n, -1, 1)$

  r2¡- x$\hat{2}$ + y$\hat{2}$; $r < -sqrt(r2)$

- Summary:

  $mean(r)$; $var(r)$; $hist(r)$;

  lab ¡- "100 samples from a uniform distribution over [-1,1] × [-1,1]"; $hist(r, main = lab)$

In one experiment, it resulted in 0.77476 and 0.08158. We can use this technique to evaluate the probability of $R \leq 0.7$.

We consider two simple techniques for computing one-dimensional integrals,

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

by a Monte Carlo mothod. The first technique is called "the hit or miss Monte Carlo method," and is based on the geometrical interpretation of an integral as an area; the second technique is called "the sample-mean Monte Carlo method," and is based on the representation of an integral as a mean value.

To motivated the need of computing one-dimensional integrals, we start with an example involving random parameters. It is common to model the number of automobile accidents in a year for a given driver in a certain population has a Poisson distribution with mean λ. Poisson distribution is found to be useful in modeling the occurrence of rare events. Here λ is referred to as the *accident proneness* of the driver. It varies from driver to driver. We can model it by a gamma distribution with shape parameter 2 and inverse-scale parameter 10.

For insurance company, it would like to know the number of accidents for all drivers. In such a case, we can set Y has the Poisson distribution with mean λ. Or,

$$f(y|\lambda) = \frac{\lambda^y}{y!} exp(-\lambda), \quad y =, 1, 2, \ldots.$$

Here λ ranges over a set Λ and is a random parameter which is associated with each individual (or object) in a certain population. As λ varies from individual to individual in the population, it has a distribution with a density function on Λ. (In practical applications, of course, the population is finite, so λ must necessarily have a discrete distribution. We assume, however, that the population size is so large that this discrete distribution can be accurately approximated by a continuous distribution.) We refer to the density function of λ as its *prior density function*. Then $f(y|\lambda)$ is the conditional density function of Y given λ. Let the prior distribution of λ be gamma with shape parameter 2 and inverse-scale parameter 10. Then the prior density function of λ is given by

$$f_{prior}(\lambda) = \frac{10^2 \lambda^{2-1}}{\Gamma(2)} exp(-10\lambda), \quad \lambda > 0.$$

The marginal probability function of Y is given by

$$f_Y(y) = \int_\Lambda f_{prior}(\lambda) \cdot f(y|\lambda) d\lambda;$$

the corresponding distribution is referred to as the *marginal distribution* of Y. The condition density function of λ, given that $Y = y$, is referred to as the posterior density function of λ which is $f_{prior}(\lambda) f(y|\lambda)/f_Y(y)$. The corresponding distribution is referred to as the *posterior distribution* of λ.

If we adopt the squared error loss, the best predictor of λ based on Y is given by $\hat{\lambda} = E(\lambda|Y)$. It is also referred to as the *Bayes estimate* of λ (based on Y). Thus, $\hat{\lambda}$ is the mean of the posterior distribution of λ; that is,

$$\hat{\lambda} = \int_{\Lambda} f_{post}(\lambda)d\lambda = \frac{\int_{\Lambda} \lambda f_{prior}(\lambda)f(Y|\lambda)d\lambda}{\int_{\Lambda} f_{prior}(\lambda)f(Y|\lambda)d\lambda}.$$

For the above example, the marginal probability function of Y is

$$f_Y(y) = \frac{10^2\Gamma(2+y)}{\Gamma(2)y!(10+1)^{2+y}}, \quad y = 0, 1, 2, \ldots,$$

the posterior distribution of λ is gamma with shape parameter $2+Y$ and inverse-scale parameter $10+1$, and $\hat{\lambda} = (2+Y)/(10+1)$. Suppose a randomly selected driver has no accidents during the year. The Bayes estimate of λ is $(2+0)/(10+1)$.

We now consider another example. The IQ of a randomly selected individual in a certain population is normally distributed with mean 100 and standard deviation 15. The score of the individual on an IQ test is unbiased and normally distributed with standard deviation 5. Note that Y is normally distributed with mean θ and variance 5 and the prior distribution of θ be normal with mean 100 and standard deviation 15. Then the posterior distribution of θ is normal with mean $(5^2 \cdot 100 + 15^2 \cdot Y)/(5^2+15^2)$ and variance $5^2 \cdot 15^2/(5^2+15^2)$ and $\hat{\theta} = 118$ if $Y = 120$.

1.1 The Hit or Miss Monte Carlo Method

Consider the problem of calculating the one-dimensional integral (??) where, for simplicity, we assume that the integrand $g(x)$ is bounded

$$0 \leq g(x) \leq c, \quad a \leq x \leq b.$$

Let Ω denote the rectangle

$$\Omega = \{(x, y) : a \leq x \leq b, 0 \leq y \leq c\}.$$

Let (X, Y) be a random vector uniformly distributed over the rectangle Ω with probability density function (p.d.f.)

$$f_{XY}(x, y) = \begin{cases} \frac{1}{c(b-a)}, & \text{if } (x, y) \in \Omega \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

What is the probability p that the random vector (X, Y) falls within the area under the curve $g(x)$? Denoting $S = \{(x, y) : y \leq g(x)\}$ and observing that the area under the curve $g(x)$ is

$$\text{area under } g(x) = \text{area } S = \int_a^b g(x)dx,$$

we obtain

$$p = \frac{\text{area } S}{\text{area } \Omega} = \frac{\int_a^b g(x)dx}{c(b-a)} = \frac{I}{c(b-a)}. \tag{3}$$

Let us assume that N independent random vectors $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$ are generated. The parameter p can be estimated by

$$\hat{p} = \frac{N_H}{N}, \tag{4}$$

where N_H is the number of occasions on which $g(X_i) > Y_i$, $i = 1, 2, \ldots, N$, that is, the number of "hits," and $N - N_H$ is the number of "misses"; we score a miss if $g(X_i) < Y_i$, $i = 1, 2, \ldots, N$.

It follows from (??), and (??) that the integral I can be estimated by

$$I \approx \theta_1 = c(b-a)\frac{N_H}{N}. \tag{5}$$

In other words, to estimate the integral I we take a sample N from the distribution (??), count the number N_H of hits (below the curve $g(x)$), and apply (??).

Since each of the N trials constitutes a Bernoulli trial with probability p of a hit, then

$$E(\theta_1) = I \tag{6}$$

$$Var(\theta_1) = \frac{I}{N}[c(b-a) - I]. \tag{7}$$

Note that the precision of the estimator θ_1, which is measured by the inverse of standard deviation, is of order $N^{-1/2}$.

How many trials do we have to perform, according to the hit or miss Monte Carlo method, if we require

$$P[|\theta_1 - I| < \epsilon] \geq \alpha? \tag{8}$$

Chebyschev's inequality,

$$P[|\theta_1 - I| < \epsilon] \geq 1 - \frac{Var(\theta_1)}{\epsilon^2},$$

together with (??), (??), gives

$$\alpha \leq 1 - \frac{p(p-1)[c(b-a)]^2}{N\epsilon^2}. \tag{9}$$

Solving (??) with respect to N, we have

$$N \geq \frac{p(1-p)[c(b-a)]^2}{(1-\alpha)\epsilon^2}$$

which is the required number of trial for (??) to hold.

When N is sufficiently large we can apply the central limit theorem, we can easily verify that the confidence interval with level $1-2\alpha$ for I is

$$\theta_1 \pm z_\alpha \frac{\hat{p}(1-\hat{p})(b-a)c}{N^{1/2}}, \tag{10}$$

where $z_\alpha = \phi^{-1}(\alpha)$.

Consider estimation of π through the representation

$$\frac{\pi}{4} = \int_0^1 \sqrt{1-x^2}dx, \tag{11}$$

each side of (??) being the area of a quadrant of a circle of radius unity. Then $Var(\theta_1) = \pi(4-\pi)/N \approx 2.697/N$.

Suppose that we want to evaluate $\int_0^1 xdx$. Again, use S statistical system, it can be done easily as follows:

- Generate Data:

  $n < -100$; $x < -runif(100)$; $y < -runif(100)$; $g < -x$

- Calculation of Missing:

  $nmiss < -pmin(1, y/g)$; $nmiss[nmiss < 1] < -0$;

- Summary:

  $sum(nhit)/100$

In one experiment, it resulted in 0.44. For the same data set, we get 0.489 for the technique discussed in next section.

1.2 Crude Monte Carlo or The Sample-Mean Monte Carlo Method

Another way of computing the integral I is to represent it as an expected value of some random variable. Indeed, let us rewrite the integral as

$$I = \int_a^b \frac{g(x)}{f_X(x)} f_X(x)dx, \tag{12}$$

assuming that $f_X(x)$ is any p.d.f. such that $f_X(x) > 0$ when $g(x) \neq 0$. Then $I = E[g(X)/f_X(X)]$ where the random variable X is distributed according to $f_X(x)$. For simplicity, we can assume that X is a uniform random variable over (a, b). Then $Eg(X) = (b-a)^{-1}I$. An unbiased estimator of I is

$$\theta_2 = (b-a)\frac{1}{N}\sum_{i=1}^{N} g(X_i).$$

The variance of θ_2 is

$$Var(\theta_2) = \frac{1}{N}\left[(b-a)\int_a^b g^2(x)dx - I^2\right].$$

As $\pi/4 = \int_0^1 \sqrt{1-x^2}\cdot 1dx$, we can write $\pi/4 = E[\sqrt{1-U^2}]$, where U is a $U(0,1)$ random variable, and so if we take a random sample, $U_1, U_2, \ldots, U_N$, we can estimate $\pi/4$ by:

$$I = N^{-1}\sum_{i=1}^{N}\sqrt{1-U_i^2}.$$

This approach is termed *crude* Monte Carlo. Clearly, $Var(\theta_2) = N^{-1}\left(\frac{2}{3} - \frac{\pi^2}{16}\right) \approx 0.0498N^{-1}$.

By now, we have two Monte Carlo methods for estimating the integral I. Let us compare now the efficiency of the hit or miss Monte Carlo method with that of the crude Monte Carlo method. To construct a point at random within a unit square, we only need to take a point with Cartesian coordinates (U_1, U_2), where U_1 and U_2 are independent $U(0,1)$ variables. In order to compare two approaches to estimating π we can take $2N$ $U(0,1)$ variables in each case, resulting in the respective variance of estimators of π: $2.697N^{-1}$ and $0.0498 \times 16/2N = 0.398N^{-1}$. So we see the variance in the hit-or-miss case is roughly seven times larger than in the crude case, indicating that crude Monte Carlo is much more efficient.

1.3 Limit Theorems

Markov's inequality. Suppose ϕ is nonnegative and nondecreasing on $[0, \infty)$ and $X \geq 0$. Then, when $\phi(y) > 0$,

$$P(X \geq y) \leq E\phi(X)/\phi(y).$$

Note that Chebyshev's inequality is a special case of Markov's inequality By taking $Y = |X - E(X)|$ and $\phi(x) = x^2$.

Use the above inequality, we can derive the following two theorems:

- **Weak law of large numbers.** Suppose $X_1, X_2, \ldots$ are i.i.d. with $E|X_1| < \infty$. Then as $n \to \infty$, $\bar{X}_n$ convereges to $E(X_1)$ in probability.

- **Strong law of large numbers.** Suppose $X_1, X_2, \ldots$ are i.i.d. with $E|X_1| < \infty$. Then with probability one the sequence of numbers $\bar{X}_n$ convereges to $E(X_1)$ as $n \to \infty$.

The weak law of large numbers can be called the fundamental theorem of statistics because it says that the sample mean is close to the mean μ of the underlying population when the sample is large. Because it does not rule out the possibility that the sequence of sample means $\bar{X}_1, \bar{X}_2, \ldots$ stays close to $E(X)$ most of the time but occasionally wanders off because of a streak of bad luck. Strong law of large numbers says that this does not happen. Next we will deal with the important question: How close will the mean of a sample of size $1,000$ be to the mean of the underlying population μ?

Chebyshev's inequality tells that the convergence rate is $n^{-1/2}$. The central limit theorem will give more information on the distance.

Central limit theorem. Suppose $X_1, X_2, \ldots$ are i.i.d. and have $EX_i = \mu$ and variance $var(X_i) = \sigma^2$ with $0 < \sigma^2 < \infty$. Then as $n \to \infty$,

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) \to P(Z \leq x)$$

Where Z stands a random variable with the standard normal distribution.

Unlike the Chebyshev's inequality, the above theorem will only give an asymptotic result on the behavior of convergence. Now we state Bernstein's inequality. Apply Markov's inequality, we have for any $t > 0$,

$$P(X \geq x) \leq e^{-tx} E exp(tX).$$

Suppose X_i are exponential(1) random variables, and $S_n = X_1 + \cdots + X_n$. Then

$$P(S_n > cn) \leq exp(-n(c - 1 - \ln c))$$

with $c > 1$. Note that $P(S_n > cn) \to 0$ exponentially fast as $n \to \infty$.

Consider the sample-mean Monte Carlo method with $X \sim UNIF(0, 1)$. Suppose g has $\int_0^1 g^2(x)dx < \infty$ and let $U_1, U_2, \ldots$ be i.i.d. uniform on $(0, 1)$. Let $I_n = (g(U_1 + \cdots + g(U_n))/n$ and $\sigma^2 = \int_0^1 g^2(x)dx - I^2$. Then

$$P(\sigma^{-1} n^{1/2}(I_n - I) \leq x) \to P(Z \leq x).$$

This is not a very fast rate of convergence since by evaluating g by the Trapezoidal rule at n equally spaced points we have (if g is twice differentiable)

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g(x_{i-1})+g(x_i)}{2}-I=O(n^{-2}).$$

However, Monte Carlo integration works even when the function is not smooth and has a convergence rate of $n^{-1/2}$ in any dimension. In contrast, with a regular grid in d dimensions, you need m^d points to achieve an accuracy of m^{-2}, or changing variables $n=m^d$, we have a convergence rate of $n^{-2/d}$.

2 Variance Reduction Techniques

Variance reduction can be viewed as a means to use known information about the problem. In fact, if nothing is known about the problem, variance reduction cannot be achieved. At the other extreme, that is, complete knowledge, the variance is equal to zero and there is no need for simulation. Variance reduction cannot be obtained from nothing; it is merely a way of not wasting information.

2.1 Antithetic Variates

This technique is due to Hammersley and Morton (1956). In this technique we seek two unbiased estimators $Y^{'}$ and $Y^{''}$ for some unknown parameter I (in our case I is the unknown integral), having strong negative correlation. Note that $(Y^{'}+Y^{''})/2$ will be an unbiased estimator of I with variance

$$Var\left[\frac{1}{2}(Y^{'}+Y^{''})\right]=\frac{1}{4}Var(Y^{'})+\frac{1}{4}Var(Y^{''})+\frac{1}{2}Cov(Y^{'},Y^{''}), \tag{13}$$

and it follows from the last equation that, if the covariance $Cov(Y^{'},Y^{''})$ is strongly negative, the method of antithetic variates can be effective in reducing the variance.

As an example, consider the integral $I=\int_0^1 g(x)dx$, which is equal to

$$I=\frac{1}{2}\int_0^1[g(x)+g(1-x)]dx. \tag{14}$$

The estimator of I is then

$$I=\frac{1}{2}(Y^{'}+Y^{''})=\frac{1}{2}\left[g(U)+g(1-U)\right]. \tag{15}$$

To estimate I we can take a sample of size N from the uniform distribution and find

$$\theta_7 = \frac{1}{2N}\sum_{i=1}^{N}\left[g(U_i) + g(1-U_i)\right]. \tag{16}$$

The time required for one computation by (??) is twice that required by the sample-mean method. Therefore the estimator θ_7 will be more effective than the estimator θ_2 with $a = 0$ and $b = 1$ only if

$$Var(\theta_7) \leq \frac{1}{2}Var(\theta_2).$$

More generally, let $I = \int_{-\infty}^{\infty} g(x)f_X(x)dx$. Then I can be estimated with

$$\theta_7 = \frac{1}{2N}\sum_{i=1}^{N}[g(X_i) + g(X_i^{'})],$$

where $X_i = F^{-1}(U_i)$, $X_i^{'} = F^{-1}(1-U_i)$, and $F_X(x)$ is the cumulative distribution function of X. The pairs X_i and $X_i^{'}$ are, of course, correlated since the same random numbers U_i, $i = 1, \ldots, N$, were used to generate both r.v.'s X_i and $X_i^{'}$.

Back to the estimation of $\pi/4$. Let

$$H = \frac{1}{2}\left[\sqrt{1-U^2} + \sqrt{1-(1-U)^2}\right],$$

where U is $U(0,1)$. $1-U$ is the "antithetic" variate to U. Then we have

$$\begin{aligned} Var(H) &= \frac{1}{2}\left[Var(\sqrt{1-U^2}) + Cov(\sqrt{1-U^2}, \sqrt{1-(1-U)^2})\right] \\ &= \frac{1}{2}\left(\frac{2}{3} - \frac{\pi^2}{16}\right) + \frac{1}{2}E\left[\sqrt{(U+1)U(U-1)(U-2)} - \frac{\pi^2}{16}\right]. \end{aligned}$$

It can be shown that

$$\begin{aligned} E\left[\sqrt{(U+1)U(U-1)(U-2)}\right] &= \frac{\pi}{4}\left\{\frac{71}{96} - 6\sum_{k=2}^{\infty}\frac{(2k-3)!(2k-1)!(12)^{-2k}}{(k-2)!(k-1)!k!(k+1)!}\right\} \\ &\approx 0.5806 \end{aligned}$$

leading to $Var(H) = 0.0052$.

Thus if $2N$ $U(0,1)$ variates were used to estimate π using this antithetic approach, the resulting estimator would have variance $0.042N^{-1}$. The crude estimator variance is just nine times larger than this, while the hit-or-miss estimator variance is roughly 64 times larger.

2.2 Importance Sampling

Let us consider the problem of estimating the multiple integral

$$I = \int g(x)dx, \quad x \in D \subset R^n. \tag{17}$$

We supppose that $g \in L^2(x)$.

The basic idea of this technique consists of concentrating the distribution of the sample points in the parts of the region D that are of most "importance" instead of spreading them out evenly. We can represent the integral (??) as

$$I = \int \frac{g(x)}{f_X(x)} f_X(x) dx = E\left[\frac{g(x)}{f_X(x)}\right]. \tag{18}$$

Here X is any random vector with p.d.f. $f_X(x)$, such that $f_X(x) > 0$ for each $x \in D \subset R^n$. The function $f_X(x)$ is called the importance sampling distribution. It is obvious from (??) that $\xi = g(X)/f_X(X)$ is an unbiased estimator of I, with the variance

$$Var(\xi) = \int g^2(x)/f_X(x) dx - I^2. \tag{19}$$

In order to estimate the integral we take a sample $X_1, \ldots, X_N$ from p.d.f. $f_X(x)$ and substitute its values in the sample-mean formula

$$\theta_3 = \frac{1}{N} \sum_{i=1}^{N} \frac{g(X_i)}{f_X(X_i)}. \tag{20}$$

We now show how to choose the distribution of the random variable X in order to minimize the variance of θ_3.

Theorem 1 *The minimum of $Var(\xi)$ is equal to*

$$Var(\xi_0) = \left(\int |g(x)| dx\right)^2 - I^2 \tag{21}$$

and occurs when the rv X is distributed with pdf

$$f_X(x) = \frac{|g(x)|}{\int |g(x)| dx}. \tag{22}$$

Proof. Formula (??) follows directly if we substitute (??) into (??). In order to prove that $Var(\xi_0) \leq Var(\xi)$ it is enough to prove that

$$\left(\int |g(x)| dx\right)^2 \leq \int \frac{g^2(x)}{f_X(x)} dx,$$

which can be obtained from Cauchy-Schwarz inequality.

Indeed,

$$\begin{aligned} \left(\int |g(x)| dx\right)^2 &= \left(\frac{|g(x)|}{[f_X(x)]^{1/2}} [f_X(x)]^{1/2} dx\right)^2 \\ &= \int \frac{g^2(x)}{f_X(x)} dx \int f_X(x) dx = \int \frac{g^2(x)}{f_X(x)} dx. \end{aligned}$$

Based on the above theorem, the optimal pdf is $f_X(x) = g(x)/I$ if $g(x) > 0$. This method is unfortunately useless, since the optimal density contains the integral $\int |g(x)|dx$, which is practically equivalent to computing I. In the case where $g(x)$ has constant sign it is precisely equivalent to calculating I. But if we already know I, we do not need Monte Carlo methods to estimate it.

Not all is lost, however. The variance can be essentially reduced if $f_X(x)$ is chosen in order to have a shape similar to that of $|g(x)|$. When choosing $f_X(x)$ in such a way we have to take into consideration the difficulties of sampling from such a pdf, especially if $|g(x)|$ is not a well behaved function.

Consider the problem of choosing the parameters of the distribution $f_X(x)$ in an optimal way. We assume that the pdf $f_X(x)$ is determined up to the vector of parameters α, that is, $f_X(x) = f_X(x, \alpha)$. For instance, if $X \sim N(\mu, \sigma^2)$, then the unknown parameters can be expected value μ and the variance σ^2. We want to chose the vector of parameters α to minimize the variance of θ_3. This is equivalent to minimize $\int g^2(x)/f_X(x, \alpha)dx$ over α. For example, consider $g(x) = (2\pi)^{-1/2} x \exp(-x^2/2)$ and $X \sim N(0, \sigma^2)$. Note that $I = E(Z) = 0$ where $Z \sim N(0, 1)$) We now show that the optimal choice of σ^2 is not 1. Observe that

$$\begin{aligned} \frac{g^2(x)}{f_X(x, \alpha)} &= \frac{(2\pi)^{-1} x^2 \exp(-x^2)}{(2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)} \\ &= \sigma\sqrt{\frac{\sigma^2}{2\sigma^2 - 1}} x^2 \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2\sigma^2-1}}} \exp\left(-\frac{x^2}{2\frac{\sigma^2}{2\sigma^2-1}}\right) \end{aligned}$$

and

$$\int \frac{g^2(x)}{f_X(x, \alpha)} dx = \sigma\sqrt{\frac{\sigma^2}{2\sigma^2 - 1}} EX^2$$

where $X \sim N(0, \sigma^2/(2\sigma^2-1))$. Since $\int \frac{g^2(x)}{f_X(x,\alpha)}dx = \sigma^4/(2\sigma^2-1)^{3/2}$, we can find the minimizer by letting $\theta = \sigma^2$ and solving

$$2\ln\theta - \frac{3}{2}\ln(2\theta - 1).$$

We get $\sigma^2 = \theta = 2$.

If we take $f_X(x, \alpha) = 1/[\pi(1 + x^2)]$, then

$$\begin{aligned} \frac{g^2(x)}{f_X(x, \alpha)} &= \frac{(2\pi)^{-1} x^2 \exp(-x^2)}{\frac{1}{\pi(1+x^2)}} \\ &= \frac{\sqrt{1/2}}{2} \frac{1}{\sqrt{1/2}} (x^2 + x^4) \exp\left(-\frac{x^2}{2 \cdot \frac{1}{2}}\right) \end{aligned}$$

and

$$\int \frac{g^2(x)}{f_X(x, \alpha)} dx = 2^{-3/2} E(X^2 + X^4)$$

where $X \sim N(0, 1/2)$. We then have

$$\frac{g^2(x)}{f_X(x, \alpha)} = 5/2^{7/2}$$

which is even smaller than $16/\sqrt{27}$ from the previous approach. An interesting question to ask is why we get such a result?

2.3 Write report on Monte Carlo experiment

An important consideration in the conduct of a Monte Carlo experiment is the items to include in the report of the study. Frequently one sees in articles in the statistical journals reports of Monte Carlo studies that fail to supply any details of the study; only the results are given. Since the results are realizations of pseudorandom processes, the reader of the article may be left wondering about the magnitude of the pseudo-standard errors and, indeed, about the effect of the pseudo nature of the process itself. It should include the descriptions of the following:

- the basic random number generators used
- the transformations to nonuniform variates
- the use of any variance reducing technique
- the supporting software and the machine(s) used.