Abstract

In this work, we consider two Monte Carlo methods for evaluating the n-dimensional integrals for bounded integrand. Statistical properties of these methods are illustrated and unified. The supported number of trials to estimate the integrals, confidence interval and the efficiency for each method were derived theoretically and assessed practically. Variance Reduction for Monte Carlo methods is discussed theoretically and explained by algorithms where four techniques are considers, namely, the Importance Sampling, the Correlated Sampling, the Partition of the region, and the Biased Estimator. The computer programs are illustrated in appendices by the run is made by using MathCAD 2001*i*.

Acknowledgement

Before every thing …

Thanks to Allah for helping me to complete my thesis

I wish to express my deep appreciation and sincere thanks to my supervisor Dr. Akram M. Al-Abood for his appreciable advices, important comments support and encouragement during the research.

It is pleasure to record my independences to the staff members of Mathematics and Computer Applications, whom gave me all facilities during my work and pursuit of my academic study.

Thanks are extended to the college of Science of Al-Nahrain University for giving me the chance to complete my postgraduate study.

Last but not least, I am deeply indebted to my family, father, mother and my brothers, Ali, Ahmed ,Hussein and my uncle Jaleel, my aunt Hufdhia, and my sister Mrs. Sally, and my friends Salam, Mohammed, Omar, Mohammed, Mahdi, Ali, Ayad, Bilal, Azad, Anas, Zaid, Noor, Sara, Semma and all my friends, and every one who helped me in my project.

Akram Al-Sabbagh

April 2008

جمهورية العراق

وزارة التعليم العالي و البحث العلمي

جامعة النهرين

كلية العلوم

تكاملات مونت كارلو و تقنيات تخفيض التباين للتكاملات المتعددة الأبعاد

رسالة

مقدمّة إلى كلية العلوم-جامعة النهرين، وهي جزء من متطلبات نيل درجة ماجستير في علوم الرياضيّات

مِنْ قِبَلْ

أكرم عبّاس جاسم الصبّاغ

(بكالوريوس علوم/جامعة النهرين،٢٠٠٥)

بإشراف

أ.م.د. أكرم محمّد العبود

نيسان ٢٠٠٨ | ربيع الأول ١٩٢٩

CHAPTER 1

SOME BASIC CONCEPTS AND DEFINITIONS

1.1 Introduction :

In this chapter we introduce some methods for generating random numbers on digital computers and their properties associated with uniform random variates where the term “random number” is used instead of uniform random number [15].

This chapter involve four sections. In section 1.2 we introduce some basic concepts and definitions concerning the distn. of random variables, while in section 1.3 we introduce some techniques for generating random numbers on digital computers. In section 1.4 we consider two important methods for generating random variates from different probability distn., namely, the Inverse Transform method, and the Acceptance-Rejection method. These two methods are discussed theoretically and supported by examples.

1.2 Some Basic Concepts and Definitions :

In this section we shall illustrate some basic concepts and definitions which are needed for simulation and Monte Carlo procedures.

Definition 1.2.1 (n-dimensional random vectors)[1]:

Given a random experiment with S.S Ω, a vector function $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ which assign to each element $\omega \in \Omega$ one and only one n-tuples vector of real numbers $\underset{\sim}{X}(\omega) = \underset{\sim}{x}$ is called an n-dimensional random vector.

The space of $\underset{\sim}{X}$ is the set of ordered n-tuples real numbers $\mathcal{A} = \{\underset{\sim}{x} : \underset{\sim}{x} = \underset{\sim}{X}(\omega), \omega \in \Omega\}$. The random vector $\underset{\sim}{X}$ is classified into two types:

(i) Discrete.

(ii) continuous.

Definition 1.2.2: A random vector $\underset{\sim}{X}$ is said to be discrete if it is defined on a countable S.S whether it is finite or infinite, otherwise, $\underset{\sim}{X}$ is called continuous random vector.

Definition 1.2.3 (Probability Density Function "p.d.f")[1]:

Let $\underset{\sim}{X}$ be an n-dimensional random vector "disc. or cont." define on S.S $\mathcal{A}$. A function $f(\underset{\sim}{x}) = f(x_1, x_2, \ldots, x_n)$ is called multivariate or joint p.d.f of $\underset{\sim}{X}$ "or distn" if $f(\underset{\sim}{x})$ satisfy the following two conditions:

i. $f(\underset{\sim}{x}) \geq 0, \forall \underset{\sim}{x} \in \mathcal{A}$.

ii. $$1 = \begin{cases} \sum\sum_{\underset{\sim}{x} \in A} \cdots \sum f(\underset{\sim}{x}), & \underset{\sim}{X} \ disc. \\ \int\int_{\underset{\sim}{x} \in A} \cdots \int f(\underset{\sim}{x}) d\underset{\sim}{x}, & \underset{\sim}{X} \ cont. \end{cases}$$

Definition 1.2.4 (Cumulative Distribution Function "c.d.f")[1]:

Let $\underset{\sim}{X}$ be an n-dimensional random vector with p.d.f $f(\underset{\sim}{x})$ defined on S.S $\mathcal{A}$, we define the c.d.f of $\underset{\sim}{X}$ "or distn.", denoted by $F(\underset{\sim}{x}) = \Pr[\underset{\sim}{X} \leq \underset{\sim}{x}]$, as:

$$F(\underset{\sim}{x}) = \begin{cases} \sum_{t_1=-\infty}^{x_1} \sum_{t_2=-\infty}^{x_2} \cdots \sum_{t_n=-\infty}^{x_n} f(t_1, t_2, \ldots, t_n), & \underset{\sim}{X} \ disc. \\ \int_{t_1=-\infty}^{x_1} \int_{t_2=-\infty}^{x_2} \cdots \int_{t_n=-\infty}^{x_n} f(t_1, t_2, \ldots, t_n) dt_1 dt_2 \ldots dt_n, & \underset{\sim}{X} \ cont. \end{cases}$$

Provided the sums or integrals converge analytically.

Where $F(\underset{\sim}{x})$ satisfy the following properties:

i. $0 \leq F(\underset{\sim}{x}) \leq 1$.

ii. $F(-\underset{\sim}{\infty}) = \underset{\sim}{0}$ and $F(\underset{\sim}{\infty}) = \underset{\sim}{1}$.

iii. $F(\underset{\sim}{x})$ is a monotonic non-decreasing function of $\underset{\sim}{x}$.

iv. $F(\underset{\sim}{x})$ is cont. function to the right at each $\underset{\sim}{x}$.

Definition 1.2.5 (Mathematical Expectation)[6]:

Let $\underset{\sim}{X}$ be an n-dimensional random vector defined on S.S $\mathcal{A}$ with p.d.f $f(\underset{\sim}{x})$, and let $u(\underset{\sim}{X})$ be any function of $\underset{\sim}{X}$, we define the mathematical expectation "or the expected value" of $u(x)$ "denoted by $E[u(x)]$" as:

$$E[u(\underset{\sim}{X})]=\begin{cases}\sum\sum\cdots\sum_{\underset{\sim}{x}\in A} u(\underset{\sim}{x})\cdot f(\underset{\sim}{x}), & \underset{\sim}{X} \ disc.\\ \int\int_{\underset{\sim}{x}\in A}\cdots\int u(\underset{\sim}{x})\cdot f(\underset{\sim}{x})d\underset{\sim}{x}, & \underset{\sim}{X} \ cont.\end{cases}$$

Provided the sums or integrals converge analytically.

In particular, for univariate case:

1. if $u(x)=x$, then $E[u(X)]$ is called the mean of r.v X "or distn." and denoted by μ.

2. if $u(x)=(x-\mu)^2$, then $E[(X-\mu)^2]$ is called the variance of the r.v X "or distn.", denoted by σ^2 or $\mathrm{var}(X)$. The positive square root of the variance σ^2 is called the standard deviation, denoted by σ .which is measure of dispersion.

In practice:

$$\sigma^2=E[(X-\mu)^2]=E[X^2-2X\mu+\mu^2]=E[X^2]-2\mu\cdot E[x]+\mu^2=E[X^2]-\mu^2$$

Moreover, for the multivariate case:

1. If $u(\underset{\sim}{x})=x_1^{r_1}\cdot x_2^{r_2}\cdot\cdots\cdot x_n^{r_n}=\prod_{i=1}^{n}x_i^{r_i}$, then

$$E\left[\prod_{i=1}^{n}x_i^{r_i}\right]=\begin{cases}\sum\sum_{\underset{\sim}{x}\in A}\cdots\sum\prod_{i=1}^{n}x_i^{r_i}\cdot f(\underset{\sim}{x}), & \underset{\sim}{X} \ disc.\\ \int\int_{\underset{\sim}{x}\in A}\cdots\int\prod_{i=1}^{n}x_i^{r_i}\cdot f(\underset{\sim}{x})d\underset{\sim}{x}, & \underset{\sim}{X} \ cont.\end{cases}$$

Which defines the multivariate moment of order $(r_1,r_2,\ldots,r_n)$ about the origin of the distn. of $\underset{\sim}{X}$.

2. If $u(\underset{\sim}{x})=\exp(t_1x_1+t_2x_2+\cdots+t_nx_n)=\exp(\sum_{i=1}^{n}t_ix_i)=e^{\underset{\sim}{t}^T\underset{\sim}{x}}$,

then

$$E[u(\underset{\sim}{X})] = \begin{cases} \sum\sum_{\underset{\sim}{x}\in A}\cdots\sum \exp(\sum_{i=1}^{n} t_i x_i)\cdot f(\underset{\sim}{x}), & \underset{\sim}{X} \ disc. \\ \int\int_{\underset{\sim}{x}\in A}\cdots\int \exp(\sum_{i=1}^{n} t_i x_i)\cdot f(\underset{\sim}{x})d\underset{\sim}{x}, & \underset{\sim}{X} \ cont. \end{cases}$$

Which defines the multivariate moment generating function, and denoted by $M(\underset{\sim}{t}) = M(t_1, t_2, \ldots, t_n)$ for the distn. of $\underset{\sim}{X}$.

Definition 1.2.6 (Statistic) [6]:

A statistic is a function of one or more r.v ^{s} which is not depend upon any unknown parameter.

Definition 1.2.7 (Estimator) [6]:

Any statistic whose values are used to estimate the unknown parameter θ or some function of θ say τ(θ) is called point estimator.

Definition 1.2.8 (Unbiased Estimator) [1]:

An estimator $U= u(X_1,X_2,\ldots,X_n)$ is said to be an unbiased estimator of θ if and only if $E[U] = E[u(X_1, X_2, \ldots, X_n)] = \theta$, denoted by $\hat{\theta}$.

The term $E[U] - \theta$ is called the bias of the estimator $\hat{\theta}$.

Definition 1.2.9 (Minimum Variance Unbiased Estimator) [1]:

Let $X_1,X_2,\ldots,X_n$ be a r.s of size n whose p.d.f $f(\underset{\sim}{x}, \theta)$, an estimator $U^*=u^*(X_1,X_2,\ldots,X_n)$ of θ is defined to be minimum variance unbiased estimator of θ if and only if:

1. $E[U^*] = \theta$, that is U^* is unbiased.

2. $\text{var}(U^*) \le \text{var}(U)$ for any unbiased estimator $U = u(X_1, X_2, \ldots, X_n)$ of θ.

1.3 Random Numbers Generation:

Many techniques for generating random numbers have been suggested, tested, and used in recent years, some of these are based on random phenomena, others on deterministic recurrence procedures.

Initially, manual methods were used, including such techniques as coin flipping, dice rolling, card shuffling, and roulette wheels, but these methods were too slow for general use, and moreover, sequences generated by them could not be reproduced.

Shortly following with the computer aid it became possible to obtain random numbers.

John Von Neumann (1951) [19] suggested the Mid-Square method, using the arithmetic operations of computer, his idea was to take the square of the preceding random number and extract the middle digits, for instance, if we wish to generating a sequence of four-digits numbers:

1. Choose any 4-digit number, say 5232.
2. Square it, we obtain 27373824.
3. The next 4-digit number is of the middle 4-digit in step 2, that is 3738.
4. repeat the prosses.

But this sequence is not really random, it only seems so, in fact referred to as pseudorandom or quasi-random; still we call it random, with the appropriate reservation. Von Neumann's method likewise proved slow and awkward for statistical analysis. In addition the sequence tend to cyclicity, and once a zero is encountered the sequence terminates.

One method of generating random numbers on a digital computer consists of preparing a table and storing it in the memory of the computer. RAND Corporation (1955) [14] published a well known table of a million random digits that may be used in forming such a table. The advantage of this method is reproducibility, and its disadvantage is its lack of speed and the risk of exhausting the table.

We say that the random numbers generated by this or any other method is good one if the random numbers are;

1. Uniformly distributed.
2. Statistically independent.
3. Reproducible.

Also a good method is necessarily fast and requires minimum memory capacity.

The congruential methods for generating pseudorandom numbers are designed specifically to satisfy as many of these requirements as possible.

Many random number generators in use today are linear congruential generators, introduced by Lehmer (1951) [9], which designed to generate sequences of pseudorandom numbers according to some recursive formula based on calculating the residues modulo of some integer m of a linear transformation.

Knuth D.E (1969) [8] show that the numbers generated by these sequences appear to be uniformly distributed, and statistically independent.

Congruential methods are based on a fundamental congruence relationship, which may be expressed as:

$$X_{i+1} = (aX_i + c)(\bmod m) \quad , i = 1,2,\ldots,n \qquad \ldots (1.1)$$

where a is the multiplier, c is the increment, and m is the modulus, where a, c, and m are nonnegative integers. The modulo notation $(\bmod m)$ means that:

$$X_{i+1} = aX_i + c - mk_i \qquad \ldots (1.2)$$

Where $k_i = [(aX_i + c)/m]$ denotes the largest positive integer in $(aX_i + c)/m$.

Given an initial starting value X_i "also called the seed", with fixed values of a,c, and m, eq.(1.2) yields a congruence relationship "modulo *m*" for any value i for the sequence $\{X_i\}$.

For example, let $a = c = X_0 = 3$ and $m = 5$, then the sequence obtained from the recursive formula $X_{i+1} = 3X_i + 3(\bmod m)$ is: $X_i = 3,2,4,0,3$.

Clearly, such a sequence will repeat itself in at most *m* steps, and will therefore be periodic,

It follow from eq.(1.2) that $X_i < m$ for all *i*. This inequality means that the period of the generator can't exceed *m*, that is, the sequence X_i contains at most *m* distinct numbers. So we must to choose *a*, *c*, and *m* as better as possible to obtain the better and largest sequence of distinct random numbers. It is noted in literatures [7, 10, 12] that good statistical results with max. periodic no. can be achieved by choosing $a = 2^7 + 1$, $c = 1$ and $m = 2^{35}$.

Generators that produce random numbers according to eq.(1.1) are called "mixed Congruential generators". The random numbers on the unit interval (0,1) can be obtained by:

$$U_i = \frac{X_i}{m} \qquad \dots (1.3)$$

We note that in present days the IBM system/360 uniform random number generator, introduce a multiplicative Congruential generator of the form $X_{i+1} = aX_i (\bmod m)$ that utilizes the full word size, which is equal to 32 bits with 1 bit resaved for algebraic sign, therefore an obvious choice for m is 2^{31}.

1.4 Random Variate Generation:

There are many techniques and several alternative algorithms for generating random variates from different distribution. However, nearly all these techniques can be classified according to their theoretical basis.

We shall utilize two methods for generating r.v's of continuous type, namely Inverse Transform method and Acceptance-Rejection method.

1.4.1 Inverse Transform Method :

Let X be a cont. r.v with cumulative distribution function "c.d.f" $F(x)$. According to the properties given in section 1.2.4 $F(x)$ is non-decreasing function. The inverse function $F^{-1}(y)$ may be defined for any value of y between 0 and 1.

The inverse transform method based on the following theorem:

Theorem(1.4.1.1) [15]:

The r.v $U = F(x) \sim U(0,1)$ if and only if the r.v $X = F^{-1}(U)$ has c.d.f $\Pr[X \le x] = F(x)$.

I.T Algorithm:

1. Generate U from $U(0,1)$.
2. Set $X = F^{-1}(u)$.
3. Deliver X as a r.v generated from the distribution whose p.d.f $f(x)$.

For illustration, we shall consider the following two examples.

Example(1.4.1.1): If we wish to generate r.v from the Weibull distribution " $X \sim W(\alpha,\beta)$ ", with p.d.f:

$$f(x)=\begin{cases}\alpha\beta\cdot x^{\alpha-1}\cdot e^{-\beta\cdot x^{\alpha}} & 0<x<\infty\\ 0 & e.w\end{cases}$$

The c.d.f of X is:

$$F(x)=\Pr[X\le x]=\int_{-\infty}^{x} f(t)dt=\int_{0}^{x}\alpha\beta\cdot t^{\alpha-1}\cdot e^{-\beta\cdot t^{\alpha}}dt=-e^{-\beta\cdot t^{\alpha}}\Big|_{0}^{x}=1-e^{-\beta\cdot x^{\alpha}}$$

$$\therefore F(x)=\Pr[X\le x]=\begin{cases}0, & x\le 0\\ 1-e^{-\beta\cdot x^{\alpha}}, & 0<x<\infty\\ 1, & x=\infty\end{cases}$$

Now, set

$$u=F(x)\Rightarrow u=1-e^{-\beta\cdot x^{\alpha}}\Rightarrow e^{-\beta\cdot x^{\alpha}}=1-u=v\Rightarrow -\beta\cdot x^{\alpha}=\ln v\Rightarrow$$

$$x^{\alpha}=\frac{-1}{\beta}\cdot\ln v\Rightarrow x=\left[\frac{-1}{\beta}\cdot\ln v\right]^{\frac{1}{\alpha}}$$

Where v have the same distn. of u

Apply I.T algorithm:

1. Read α,β .
2. Generate U from $U(0,1)$.
3. Set $X=\left[\frac{-1}{\beta}\cdot\ln U\right]^{\frac{1}{\alpha}}$.
4. Deliver X as a r.v generated from $W(\alpha,\beta)$ distribution.

Example(1.4.1.2): Let $X_1,X_2,\ldots,X_n$ be a r.s of size n from the distn. whose p.d.f $f(x)$ and c.d.f. $F(x)$. Suppose we wish to generate Y_1 and Y_n

where $Y_1 = min(X_1,X_2,\ldots,X_n)$ and $Y_n = max(X_1,X_2,\ldots,X_n)$. From order statistics theory, the c.d.f of the 1st order statistics Y_1 is:

$$G_1(y_1) = \Pr[Y_1 \le y_1] = 1-[1-F(y_1)]^n$$

Set $u = G_1(y_1) \Rightarrow u = 1-[1-F(y_1)]^n \Rightarrow [1-F(y_1)]^n = 1-u = u \Rightarrow$

$$1-F(y_1) = u^{1/n} \Rightarrow F(y_1) = 1-u^{1/n} \Rightarrow y_1 = F^{-1}(1-u^{1/n})$$

and the c.d.f of the n^{th} order statistics Y_n is:

$$G_n(y_n) = \Pr[Y_n \le y_n] = [F(y_n)]^n$$

Set

$$u = G_n(y_n) \Rightarrow u = [F(y_n)]^n \Rightarrow F(y_n) = u^{1/n} \Rightarrow y_n = F^{-1}(u^{1/n})$$

Apply I.T algorithm:

1. Read n.
2. Generate U from $U(0,1)$.
3. Set $Y_1 = F^{-1}(1-U^{1/n})$ and $Y_2 = F^{-1}(U^{1/n})$.
4. Deliver Y_1 and Y_n as the 1^{st} and the n^{th} order statistics generated from the distn. whose p.d.f $f(x)$.

We note that to apply the inverse transform method, the c.d.f $F(x)$ must be exist in a form for which the corresponding inverse transform can be founded analytically, For example:

1. $X \sim Exp(\lambda)$ where $f(x) = \lambda^{-1} \cdot e^{-\lambda^{-1}x}, 0 < x < \infty$. (possible)
2. $X \sim G(2,1)$ where $f(x) = x \cdot e^{-x}, 0 < x < \infty$. (difficult)
3. $X \sim N(0,1)$ where $f(x) = \dfrac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}$. (impossible)

1.4.2 Acceptance – Rejection Method:

This method is dates back at least to Von Newman (1951) [19], and consists of sampling a random variate from an appropriate distribution and subjecting it to a test to determine whether or not it will be acceptable for use.

To carry out the method we represent $f(x)$ of the generated r.v X as: $f(x) = c \cdot h(x) \cdot g(x)$, where $c \geq 1$, $h(x)$ is also p.d.f, and $0 < g(x) \leq 1$. Then we generate two r.v^s, $U \sim U(0,1)$, and Y from $h(y)$, and test to see whether or not the inequality $U \leq g(Y)$ holds:

1. If the inequality holds, then accept $Y=X$ as a r.v generated from $f(x)$.
2. If the inequality violated, then reject the pair U, Y and try again.

This method is based on the following theorem:

1.4.2.1 Theorem [15]:

Let the p.d.f of r.v X represented as $f(x) = c \cdot h(x) \cdot g(x)$, where $c \geq 1$, $h(x)$ is also p.d.f, and $0 < g(x) \leq 1$.

Let U and Y be distributed $U(0,1)$ and $h(y)$, respectively, then $\Pr[Y = x \mid U \leq g(Y)] = f(x)$

A-R Algorithm:

1. Generate U from $U(0,1)$.
2. Generate Y from the p.d.f $h(y)$.
3. If $U \leq g(Y)$, Deliver X as a r.v generated from $f(x)$.
4. Go to step 1.

Note: For this method to be of practical interest the following criteria must be used in selected $h(x)$:

1. It should be easy to generate a r.v from $h(x)$.
2. The efficiency "probability" of the procedure $1/c$ should be large, that is, c should be close to one "which accurse when $h(x)$ is similar to $f(x)$ in shape".

Now, to illustrate this method, we choose c such that $f(x) \le c \cdot h(x) = \phi(x), \forall x \in I$, where $c \ge 1$.

The problem then is to find a function $\phi(x)$ and a function $h(x) = \phi(x)/c$, from which the r.v's can be easily generated

The maximum efficiency is achieved when $f(x) = \phi(x), \forall x \in I$. In this case $1/c = c = 1$, $g(x) = 1$.

There exist an infinite numbers of ways to choose $h(x)$ to satisfy $f(x) = c \cdot h(x) \cdot g(x)$.

For illustration, we shall consider the following two examples.

Example(1.4.2.1): if we wish to generate r.v from the distn. whose p.d.f:

$$f(x) = \frac{2}{\pi R^2}\sqrt{R^2 - x^2} \qquad , -R \le x \le R$$

We have

$$\sqrt{R^2 - x^2} \le R \qquad , \forall\, x \in [-R, R],$$

$$\frac{2}{\pi R^2}\sqrt{R^2 - x^2} \le \frac{2}{\pi R} \Rightarrow f(x) \le \frac{2}{\pi R} = \phi(x).$$

Now : $c \cdot h(x) = \phi(x) \Rightarrow \int_{-R}^{R} c \cdot h(x)dx = \int_{-R}^{R} \phi(x)dx \Rightarrow c = \int_{-R}^{R} \frac{2}{\pi R}dx = \frac{4}{\pi}$

$$h(x) = \frac{\phi(x)}{c} = \frac{2/\pi R}{4/\pi} = \frac{1}{2R}$$

and $H(x) = \int_{-R}^{x} h(t)dt = \int_{-R}^{x} \frac{1}{2R}dt = \frac{x-R}{2R}$

set $u_2 = H(y) \Rightarrow u_2 = \frac{y-R}{2R} \Rightarrow y = (2u_2 - 1)R$

and

$$g(y) = \frac{f(y)}{\phi(y)} = \frac{\frac{2}{\pi R^2}\sqrt{R^2 - x^2}}{\frac{2}{\pi R}} = \frac{1}{R}\sqrt{R^2 - x^2} = \frac{1}{R}\sqrt{R^2 - R^2(2u_2 - 1)^2}$$

$$= \sqrt{1-(2u_2-1)^2}$$

Apply AR-Algorithm:

1. Read R.
2. Generate U_1 and U_2 from $U(0,1)$.
3. Set $Y=(2U_2-1)R$.
4. If $U_1 \leq g(Y) = \sqrt{1-(2U_2-1)^2}$, deliver "we accept" $Y=X$ as a r.v generated from $f(x)$.

Go to step 2.

Example(1.4.2.2): If we wish to generate r.v from beta distribution "$X \sim Be(\alpha, \beta)$" with p.d.f:

$$f(x)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot x^{\alpha-1}\cdot(1-x)^{\beta-1} & ,0<x<1\\ 0 & ,e.w\end{cases}$$

If we choose $h(x)=\alpha\cdot x^{\alpha-1}$, $0<x<1$, and $g(x)=(1-x)^{\beta-1}$, in this case $c=\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}$,

So

$$H(x)=\int_0^x h(t)dt=\int_0^x \frac{1}{\alpha}\cdot t^{\alpha-1}dt=x^{\alpha}$$

Set $u_2=H(y)\Rightarrow u_2=y^{\alpha}\Rightarrow y=(u_2)^{\frac{1}{\alpha}}$, and $g(y)=(1-y)^{\beta-1}$

Apply A-R algorithm:

1. read α and β.
2. Generate U_1 and U_2 from $U(0,1)$.
3. Set $Y=U_2^{1/\alpha}$.
4. If $U_1\le g(Y)=(1-Y)^{\beta-1}$, Deliver "we accept" $Y=X$ as a r.v generated from $Be(\alpha,\beta)$.
5. Go to step 2.

CHAPTER 2

MONTE CARLO INTEGRATION METHODS

2.1 Introduction:

The importance of good numerical integration schemes is evident. There are many deterministic quadrature formulas for computation of ordinary integrals with well behaved integrands such as trapezoidal, Simpson's, and Gauss quadrature rules. But these numerical techniques become less attractive if the function fail to be regular "i.e. to have continuous derivatives of moderate order", especially in the case of multidimensional integrals where application of such rules runs into severs difficulties. It is often more convenient to compute such integrals by Monte Carlo methods, which, although less accurate than conventional quadratures rules, but it is much simpler to use.

This chapter involve tow sections, in section 2.2, we consider two techniques for computing the n-dimensional integrals namely the hit or miss Monte Carlo method, and the sample mean Monte Carlo method, where these two methods are supported by examples, Chebyshev's inequality is used to evaluate the number of trails to perform according to hit or miss method as well as the confidence interval for the estimated integral is derived, efficiencies of the two methods are discussed.

2.2 Monte Carlo Integration for n-dimensional integrals:

For computing n-dimensional integrals,

$$I = \int_{a_1}^{b_1}\int_{a_2}^{b_2}...\int_{a_n}^{b_n} g(x_1,x_2,...,x_n)\,dx_1 dx_2 ...dx_n \qquad ...(2.1)$$

We shall consider two techniques, the 1st is called "the hit or miss Monte Carlo method", which is based on the geometrical interpretation of an integrals as hyper volume under the surface of the integrand $g(x_1,x_2,...,x_n)$. The 2nd is called "the sample mean Monte Carlo method", which is based on the representation of an integral as an expected value, and the problem of estimating an integral by Monte Carlo method is equivalent to the problem of estimating an unknown parameter.

2.2.1 Hit or Miss Monte Carlo method:

Consider the problem of calculating the n-dimensional integral of eq.(2.1), where, for simplicity, we assume that the integrand $g(x_1,x_2,...,x_n)$ is bounded

$$0 \le g(x_1,x_2,...,x_n) \le c,\quad a_i \le x_i \le b_i,\quad i=1,2,...,n$$

Let $(X_1,X_2,\ldots,X_n,Y)$ be a random vector uniformly distributed over the region $\Omega = \{(x_1,x_2,...,x_n,y): a_i \le x_i \le b_i,\ \ 0 \le y \le c,\ \ i=1,2,...,n\}$,

with p.d.f

$$f(x_1,x_2,...,x_n,y) = \frac{1}{c(b_1-a_1)(b_2-a_2)...(b_n-a_n)},\quad (x_1,x_2,...,x_n) \in \Omega \qquad ...(2.2)$$
$$= 0,\quad e.w$$

Let p be the probability that the random vector $(X_1,X_2,\dots,X_n,Y)$ is falls within the hyper-volume under $g(x_1,x_2,\dots,x_n)$, denoted by,

$V = \{(x_1,x_2,\dots,x_n): y \le g(x_1,x_2,\dots,x_n)\}$ and observing that the hyper-volume under $g(x_1,x_2,\dots,x_n)$,

$$hyper-volume\,V = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\dots\int_{a_n}^{b_n} g(x_1,x_2,\dots,x_n)\,dx_1 dx_2 \dots dx_n$$

We obtain:

$$p = \frac{hyper-volume\,V}{hyper-volume\,\Omega} = \frac{\int_{a_1}^{b_1}\int_{a_2}^{b_2}\dots\int_{a_n}^{b_n} g(x_1,x_2,\dots,x_n)\,dx_1 dx_2 \dots dx_n}{(b_1-a_1)(b_2-a_2)\dots(b_n-a_n)}$$

$$= \frac{I}{c\,(b_1-a_1)(b_2-a_2)\dots(b_n-a_n)} \qquad \dots(2.3)$$

Let us assume that N independent random vectors $(X_{11},X_{21},\dots,X_{n1},Y_1)$, $(X_{12},X_{22},\dots,X_{n2},Y_2)$, $\dots,(X_{1N},X_{2N},\dots,X_{nN},Y_N)$ are generated. The parameter p can be estimated by

$$\hat{p} = \frac{N_H}{N} \qquad \dots(2.4)$$

Where N_H is the number of occasions on which $g(x_{1i},x_{2i},\dots,x_{ni}) \ge y_i$, $i = 1,2,\dots,N$ that is, the no. of "hits", and N-N_H is the no. of "misses", we score a miss if $g(x_{1i},x_{2i},\dots,x_{ni}) < y_i$, $i = 1,2,\dots,N$. It follows from eq.(2.3), and eq.(2.4), that the integral I can be estimated by

$$I \approx \hat{\theta}_1 = c\,(b_1-a_1)(b_2-a_2)\dots(b_n-a_n)\,\frac{N_H}{N} \qquad \dots(2.5)$$

In other wards, to estimate the integral I we take sample N from the distn. eq.(2.2), count the no. N_H of hits "below the hyper-surface $g(x_1,x_2,\dots,x_n)$", and apply eq.(2.5).

The necessary steps to estimating the integral of eq.(2.5) by hit or miss Monte Carlo method can be describe by HM3-Algorithm:

HM-Algorithm:

1. Generate a seq. $\{U_i\}_{i=1}^{N(n+1)}$ of $(n+1)N$ random numbers.
2. Arrange the random numbers into N pairs, $(U_{11}, U_{21}, ..., U_{(n+1)1})$, $(U_{12}, U_{22}, ..., U_{(n+1)2}), ..., (U_{1N}, U_{2N}, ..., U_{(n+1)N})$ in any fashion s.t each random no. U_i is used exactly once.
3. Compute
   $X_{1i} = a_1 + (b_1 - a_1)U_{1i}$, $X_{2i} = a_2 + (b_2 - a_2)U_{2i}$, $..., X_{Ni} = a_N + (b_N - a_N)U_{Ni}$
   and $g(X_{1i}, X_{2i}, ..., X_{ni}),\ \forall i = 1, 2, ..., N$.
4. Count the no. of cases N_H for which $g(X_{1i}, X_{2i}, ..., X_{ni}) > c\,U_{(n+1)i}$.
5. Estimate the integral I by $\hat{\theta}_i = c(b_1 - a_1)(b_2 - a_2)...(b_n - a_n)\dfrac{N_H}{N}$.

2.2.1.1 Statistical properties of the estimator $\hat{\theta}_1$:

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

$$E(\hat{\theta}_1) = c(b_1 - a_1)(b_2 - a_2)...(b_n - a_n)E\left[\frac{N_H}{N}\right]$$

$$= c(b_1 - a_1)(b_2 - a_2)...(b_n - a_n)\frac{1}{N}E(N_H)$$

$$E(\hat{\theta}_1) = p\,c(b_1 - a_1)(b_2 - a_2)...(b_n - a_n) = I \qquad ...(2.6)$$

That is, $\hat{\theta}_1$ is an unbiased estimator of I.

The variance of $\hat{\theta}_1$ is:

$$\text{var}(\hat{\theta}) = \text{var}\left[c(b_1 - a_1)(b_2 - a_2)...(b_n - a_n)\hat{p}\right]$$

$$\mathrm{var}(\hat{\theta}) = [c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)]^2 \,\mathrm{var}\left[\frac{N_H}{N}\right]$$

$$= \frac{1}{N^2}[c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)]^2 \,\mathrm{var}(N_H)$$

$$= \frac{1}{N}[c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)]^2 \,p\,(1-p)$$

$$= \frac{[c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)]^2}{N} \cdot \frac{I\,[c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n) - I\,]}{[c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)]^2}$$

$$= \frac{I}{N}[c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n) - I\,] \qquad \ldots(2.7)$$

and the standard deviation

$$\sigma_{\hat{\theta}_2} = \left[\mathrm{var}(\hat{\theta}_1)\right]^{\frac{1}{2}} = \frac{1}{\sqrt{N}}\left[I\,\{c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n) - I\,\}\right]^{\frac{1}{2}} \qquad \ldots(2.8)$$

Note that the precision of the estimator $\hat{\theta}_1$, which is the measured by the inverse of standard deviation, is of order $N^{\frac{-1}{2}}$.

2.2.1.2 Estimate the number of trials:

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

$$\Pr\left[\left|\hat{\theta}_1 - I\right| < \varepsilon\right] \geq \alpha \qquad \ldots(2.9)$$

Chebyshev's inequality,

$$\Pr\left[\left|\hat{\theta}_1 - I\right| < \varepsilon\right] \geq 1 - \frac{\mathrm{var}(\hat{\theta}_1)}{\varepsilon^2} \qquad \ldots(2.10)$$

together with eq.(2.9), gives

$$\alpha \leq 1 - \frac{\mathrm{var}(\hat{\theta}_1)}{\varepsilon^2} \qquad \ldots(2.11)$$

substituting eq.(2.7) in eq.(2.11), we obtain

$$\alpha \le 1 - \frac{p\,(1-p)\left[c\,(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)\right]^2}{N\,\varepsilon^2} \qquad \ldots(2.12)$$

by solving eq.(2.12) w.r.t N, we have

$$N \ge \frac{p\,(1-p)\left[c\,(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)\right]^2}{(1-\alpha)\,\varepsilon^2} \qquad \ldots(2.13)$$

Which is the required number of trials for eq.(2.9) to hold.

For illustration, we shall consider the following two examples, by taking small (α) say 0.01, 0.05, and small (ε) say 0.001, 0.005, and a large (p) say 0.99, 0.995 to get best result.

2.2.1.3 The (1- α)100% confidence interval for the integral I estimated by the hit or miss method:

For the large samples size, taken from non-Normal distn. "disc. or cont.", we can find with help of the Central Limit theorem, an approximate C.I for I because most distn. has limiting Normal distn. ($n \to \infty$).

Let $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$ be a r.s from a distn. that has mean I and variance σ^2 with existence of m.g.f $M(t)$, and suppose we required a C.I for I with probability $1\text{-}\alpha$ for a small α and unknown σ^2 :

According to C.L.T, the r.v $Y = \dfrac{\sqrt{n}\,(\bar{\hat{\theta}} - I)}{\sigma} \overset{app}{\sim} N(0,1)$,

Since, $S^2 \underset{sto.}{\overset{conv.}{\longrightarrow}} \sigma^2 \Rightarrow \dfrac{S^2}{\sigma^2} \underset{sto.}{\overset{conv.}{\longrightarrow}} 1 \Rightarrow v = \dfrac{S^2}{\sigma^2} \underset{sto.}{\overset{conv.}{\longrightarrow}} 1$,

Then the r.v $Z = \dfrac{Y}{v}$ has a limiting distn. As Y, that is:

$$Z = \frac{\sqrt{n}\,(\bar{\hat{\theta}} - I)/\sigma}{S/\sigma} = \frac{\sqrt{n}\,(\bar{\hat{\theta}} - I)}{S} \overset{app.}{\sim} N(0,1).$$

So, we can find from $N(0,1)$ table two no.[s], $\pm z_{1-\frac{\alpha}{2}}$, s.t:

$$\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha .$$

Now, the event

$$-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}} \equiv -z_{1-\frac{\alpha}{2}} < \frac{\sqrt{n}\,(\bar{\hat{\theta}} - I)}{S} < z_{1-\frac{\alpha}{2}}$$

$$\equiv -\frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < (\bar{\hat{\theta}} - I) < \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \equiv -\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < -I < -\bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

$$\equiv \bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < I < \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}} .$$

Therefore the approximate 100(1-α)% C.I for the integral I is:

$$\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \, , \, \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right).$$

For illustration we shall solve the following examples.

Example (2.2.1.1): Calculating the 99.5% C.I for the integration

$$I = \int_0^1 e^{-x^2}\, dx .$$

Solution:

For the best no. of trails N with $p = 0.99, \varepsilon = 0.001, c = 1$:

$$N \geq \frac{p\,(1-p)\left[c\,(b-a)\right]}{(1-\alpha)\,\varepsilon^2} = 4\times 10^4$$

By calculating $\hat{\theta}_1$, according to the Hit or Miss Monte Carlo method with number of repetition n=25, and the result are tabulated in table(2.1):

Table(2.1)

"The Estimators of $I = \int_0^1 e^{-x^2}\,dx$, using the Hit or Miss Method"

n	$\hat{\theta}_1$	**n**	$\hat{\theta}_1$	**n**	$\hat{\theta}_1$	**n**	$\hat{\theta}_1$	**n**	$\hat{\theta}_1$
1	0.737	**6**	0.737	**11**	0.741	**16**	0.743	**21**	0.744
2	0.752	**7**	0.749	**12**	0.754	**17**	0.739	**22**	0.749
3	0.753	**8**	0.743	**13**	0.744	**18**	0.745	**23**	0.745
4	0.741	**9**	0.747	**14**	0.753	**19**	0.750	**24**	0.747
5	0.748	**10**	0.744	**15**	0.746	**20**	0.740	**25**	0.747

then we can find

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i = \frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i = 0.746$$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]^2 = \frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i - 0.746\right]^2 = 2.343\times10^{-5}$

then $S = 4.84\times10^{-3}$

now, we can find from the standard normal distn.'s table two no.s $\pm z_{1-\frac{\alpha}{2}}$, s.t

$$\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

and the 99.5% C.I for the integral I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995}\ ,\ \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995}\right)$

$$\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} = 0.746 - \frac{4.84\times10^{-3}}{5}\cdot 2.6 = 0.743$$

$$\bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995} = 0.746 + \frac{4.84\times10^{-3}}{5}\cdot 2.6 = 0.749$$

Therefore the 99.5% C.I for I is: $(0.743\ ,\ 0.749)$

Example (2.2.1.2): Calculation the 99% C.I for the integral

$$I = \int_0^1\int_0^1\int_0^1 \frac{\sin(x_1 + x_2)}{\cos(x_3)} dx_3 dx_2 dx_1 .$$

Solution:

For the best no. of trails N with $p = 0.99,\ \varepsilon = 0.01$:

$$N \geq \frac{p(1-p)\left[c(b_1 - a_1)(b_2 - a_2)(b_3 - a_3)\right]}{(1-\alpha)\varepsilon^2} = 2.75 \times 10^3$$

By calculating $\hat{\theta}_1$, according to the Hit or Miss Monte Carlo method with number of repetition n=25, and the result are tabulated in table(2.2):

Table(2.2)

"the estimators of $I = \int_0^1\int_0^1\int_0^1 \frac{\sin(x_1 + x_2)}{\cos(x_3)} dx_3 dx_2 dx_1$,using Hit or Miss"

n	$\hat{\theta}_1$	**n**	$\hat{\theta}_1$	**n**	$\hat{\theta}_1$	**n**	$\hat{\theta}_1$	**n**	$\hat{\theta}_1$
1	0.935	**6**	0.931	**11**	0.940	**16**	0.971	**21**	0.960
2	0.956	**7**	0.954	**12**	0.958	**17**	0.947	**22**	0.936
3	0.944	**8**	0.937	**13**	0.949	**18**	0.975	**23**	0.975
4	0.967	**9**	0.938	**14**	0.956	**19**	0.904	**24**	0.958
5	0.949	**10**	0.965	**15**	0.938	**20**	0.973	**25**	0.927

then we can find

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i = \frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i = 0.950$$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]^2 = \frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i - 0.950\right]^2 = 2.950 \times 10^{-4}$

then $S = 0.017$

now, we can find from the standard normal distn.'s table two no.s $\pm z_{1-\frac{\alpha}{2}}$,

s.t: $\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$

and the 95% C.I for the integral I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.95} \, , \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.95}\right)$

$$\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.95} = 0.950 - \frac{0.017}{5} \cdot 1.645 = 0.944$$

$$\bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.95} = 0.950 + \frac{0.017}{5} \cdot 1.645 = 0.955$$

Therefore the 95% C.I for I is: $(0.944 \, , 0.955)$.

Example (2.2.1.3): Calculation the 99.5 % C.I for the integral

$$I = \int_0^1 \int_0^1 \int_1^2 \int_0^2 \int_1^3 e^{x_1^2 + x_2^2} + 2x_3 - \sin(x_4 + x_5) \, dx_1 dx_2 dx_3 dx_4 dx_5 .$$

Solution:

For the best no. of trails N with $p = 0.99, \varepsilon = 0.01$:

$$N \geq \frac{p\,(1-p)\left[c\,(b_1 - a_1)(b_2 - a_2)\ldots(b_5 - a_5)\right]}{(1-\alpha)\,\varepsilon^2} = 1 \times 10^4$$

By calculating $\hat{\theta}_1$, according to the Hit or Miss Monte Carlo method with number of repetition n=25, and the result are tabulated in table(2.3):

Table(2.3)

"the Estimators of

$$I = \int_0^1\int_0^1\int_1^2\int_0^2\int_1^3 e^{x_1^2+x_2^2} + 2x_3 - \sin(x_4+x_5)\, dx_1 dx_2 dx_3 dx_4 dx_5 ,$$

using Hit or Miss Method"

n	$\hat{\theta}_1$	n	$\hat{\theta}_1$	n	$\hat{\theta}_1$	n	$\hat{\theta}_1$	n	$\hat{\theta}_1$
1	19.960	6	20.477	11	20.381	16	20.452	21	20.000
2	20.050	7	20.516	12	20.615	17	19.932	22	20.205
3	20.460	8	20.213	13	19.964	18	20.210	23	20.157
4	20.170	9	20.120	14	20.170	19	20.336	24	20.150
5	19.970	10	20.610	15	20.350	20	20.270	25	20.370

then we can find

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i = \frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i = 20.244$$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]^2 = \frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i - 20.244\right]^2 = 42.5\times 10^{-3}$

then $S = 0.206$

now, we can find from the standard normal distn.'s table two no.s $\pm z_{1-\frac{\alpha}{2}}$,

s.t: $\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$

and the 99.5% C.I for the integral I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} \, , \, \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995}\right)$

$$\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} = 20.244 - \frac{0.206}{5}\cdot 2.6 = 20.137$$

$$\bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995} = 20.244 + \frac{0.206}{5}\cdot 2.6 = 20.352$$

Therefore the 99.5% C.I for I is: $(20.137 \, , \, 20.352)$

2.2.2 Sample Mean Monte Carlo method:

Another way of computing the integral

$$I=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} g(x_1,x_2,....,x_n)\,dx_1\,dx_2..dx_n$$

by represent I as an expected value of some r.v.

Indeed, let us rewrite the integral I as

$$I=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} \frac{g(x_1,x_2,...,x_n)}{f(x_1,x_2,...,x_n)} f(x_1,x_2,...,x_n)\,dx_1\,dx_2..dx_n \qquad \ldots(2.14)$$

where $f(x_1,x_2,\ldots,x_n)$ is any p.d.f, s.t $f(x_1,x_2,\ldots,x_n)>0,$ when $g(x_1,x_2,....,x_n)\neq 0$, then

$$I=E\left[\frac{g(X_1,X_2,....,X_n)}{f(X_1,X_2,....,X_n)}\right] \qquad \ldots(2.15)$$

where the r.v $X_1,X_2,\ldots,X_n$ are independent r.v^s are distributed according to $f_1(x_1), f_2(x_2),\ldots, f_n(x_n)$ where :

$$f_1(x_1)\cdot f_2(x_2)\cdot\ldots\cdot f_n(x_n)=f(x_1,x_2,...,x_n)$$

for simplicity, let us assume

$$f(x_1,x_2,...,x_n)=\begin{cases}\dfrac{1}{(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)} & ,a_i<x_i<b_i\ ,\forall i=1,2,...,n \\ =0 & ,ew\end{cases} \qquad \ldots(2.16)$$

then

$$E\left[g(X_1,X_2,...,X_n)\right]=\frac{I}{(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)} \Rightarrow$$

$$I=(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)E\left[g(x_1,x_2,...,x_n)\right]$$

we can estimate I by

$$\hat{\theta}_2=\frac{1}{N}(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)\sum_{i=1}^{N} g(X_{1i},X_{2i},...,X_{ni}) \qquad \ldots(2.17)$$

The necessary steps to estimating the integral of eq.(2.47) by sample mean Monte Carlo method can be describe by SM3-Algorithm:

SM-Algorithm:

1. Generate a seq. $\{U_i\}_{i=1}^{n \cdot N}$ of $n.N$ random numbers.
2. Arrange the random numbers into N pairs $(U_{11}, U_{21}, \ldots, U_{n1})$, $(U_{12}, U_{22}, \ldots, U_{n2})$, …, $(U_{1N}, U_{2N}, \ldots, U_{nN})$ in any fashion s.t each random number U_i is used exactly once.
3. Compute
   $X_{1i} = a_1 + (b_1 - a_1)U_{1i}$, $X_{2i} = a_2 + (b_2 - a_2)U_{2i}$, ..., $X_{ni} = a_n + (b_n - a_n)U_{ni}$; and $g(X_{1i}, X_{2i}, \ldots, X_{ni})$, $\forall i = 1, 2, \ldots, N$.
4. Compute the sample mean θ_2 according to:

   $\hat{\theta}_2 = (b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)\frac{1}{N}\sum_{i=1}^{N} g(X_{1i}, X_{2i}, \ldots, X_{ni})$, which

   estimates I.

2.2.2.1 Statistical properties of the estimator $\hat{\theta}_2$:

We can show that $\hat{\theta}_2$ is an unbiased estimator

$$E(\hat{\theta}_2) = E\left[\frac{1}{N}(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)\sum_{i=1}^{N} g(X_{1i}, X_{2i}, \ldots, X_{ni})\right]$$

$$= \frac{1}{N}(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)E\left[\sum_{i=1}^{N} g(X_{1i}, X_{2i}, \ldots, X_{ni})\right]$$

$$= \frac{1}{N}(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)\sum_{i=1}^{N} E\left[g(X_{1i}, X_{2i}, \ldots, X_{ni})\right]$$

$$= \frac{1}{N}(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)\sum_{i=1}^{N}\left[\frac{I}{(b_1 - a_1)(b_2 - a_2)\ldots(b_n - a_n)}\right]$$

$$\therefore E(\hat{\theta}_2) = \frac{1}{N}(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)\frac{N\,I}{(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)} = I$$

The variance of $\hat{\theta}_2$ is equal to $E\left(\hat{\theta}_2^2\right) - \left[E\left(\hat{\theta}_2\right)\right]^2$

$$\operatorname{var}(\hat{\theta}_2) = \operatorname{var}\left[\frac{1}{N}(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)\sum_{i=1}^{N} g(X_{1i}, X_{2i}, \ldots, X_{ni})\right]$$

$$= \frac{1}{N}\left[(b_1-a_1)^2(b_2-a_2)^2\ldots(b_n-a_n)^2 \int_{a_1}^{b_1}\int_{a_2}^{b_2}\ldots\int_{a_n}^{b_n} \frac{g^2(X_{1i}, X_{2i}, \ldots, X_{ni})}{(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)}\, dx_1 dx_2 \ldots dx_n - I^2\right]$$

$$\operatorname{var}(\hat{\theta}_2) = \frac{1}{N}\left[(b_1-a_1)(b_2-a_2)\ldots(b_n-a_n)\int_{a_1}^{b_1}\int_{a_2}^{b_2}\ldots\int_{a_n}^{b_n} g^2(X_{1i}, X_{2i}, \ldots, X_{ni})\, dx_1 dx_2 \ldots dx_n - I^2\right]. \quad \ldots(2.18)$$

2.2.2.2 The (1- α)100% confidence interval for the integration I estimated by the sample mean method:

From sec.(2.2.1.3) we can obtain that the approximation (1-α)100% C.I for I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\, ,\, \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$.

For illustration we will solve the following examples.

Example (2.2.2.1): Calculating the 99.5% C.I for the integral

$$I = \int_0^1 e^{-x^2}\, dx\,.$$

Solution:

For the best no. of trails N with $p = 0.99,\ \varepsilon = 0.001$:

$$N \geq \frac{p\,(1-p)\left[c\,(b-a)\right]}{(1-\alpha)\,\varepsilon^2} = 4\times 10^4$$

By calculating $\hat{\theta}_2$, according to the Sample Mean Monte Carlo method with number of repetition n=25, and the result are tabulated in table(2.4):

Table(2.4)

"The Estimators of $I = \int_0^1 e^{-x^2}\,dx$, using Sample Mean Method"

n	$\hat{\theta}_2$	**n**	$\hat{\theta}_2$	**n**	$\hat{\theta}_2$	**n**	$\hat{\theta}_2$	**n**	$\hat{\theta}_2$
1	0.744	**6**	0.744	**11**	0.750	**16**	0.751	**21**	0.748
2	0.746	**7**	0.745	**12**	0.749	**17**	0.750	**22**	0.745
3	0.747	**8**	0.747	**13**	0.749	**18**	0.746	**23**	0.746
4	0.746	**9**	0.750	**14**	0.745	**19**	0.748	**24**	0.749
5	0.746	**10**	0.747	**15**	0.748	**20**	0.744	**25**	0.748

then we can find

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i = \frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i = 0.747$$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]^2 = \frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i - 0.746\right]^2 = 4.277\times 10^{-6}$

then $S = 2.068\times 10^{-3}$

now, we can find from the standard normal distn.'s table two no.s $\pm z_{1-\frac{\alpha}{2}}$,

s.t: $\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$

and the 99.5% C.I for the integral I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}}z_{0.995}\ ,\ \bar{\hat{\theta}} + \frac{S}{\sqrt{n}}z_{0.995}\right)$

$$\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} = 0.747 - \frac{2.068 \times 10^{-3}}{5} \cdot 2.6 = 0.746$$

$$\bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995} = 0.747 + \frac{2.068 \times 10^{-3}}{5} \cdot 2.6 = 0.748$$

therefore the 99.5% C.I for I is: $(0.746\ ,\ 0.748)$

Example (2.2.2.2): Calculation the 99% C.I for the integral

$$I = \int_0^1 \int_0^1 \int_0^1 \frac{\sin(x_1 + x_2)}{\cos(x_3)} dx_3 dx_2 dx_1 .$$

Solution:

For the best no. of trails N with $p = 0.99$, $\varepsilon = 0.01$:

$$N \geq \frac{p(1-p)\left[c\,(b_1 - a_1)(b_2 - a_2)(b_3 - a_3)\right]}{(1-\alpha)\,\varepsilon^2} = 2.75 \times 10^3$$

By calculating $\hat{\theta}_2$, according to the Sample Mean Monte Carlo method with number of repetition n=25, and the result are tabulated in table(2.5):

Table(2.5)

"The Estimators of $I = \int_0^1 \int_0^1 \int_0^1 \frac{\sin(x_1 + x_2)}{\cos(x_3)} dx_3 dx_2 dx_1$, using Sample Mean method"

n	$\hat{\theta}_2$	n	$\hat{\theta}_2$	n	$\hat{\theta}_2$	n	$\hat{\theta}_2$	n	$\hat{\theta}_2$
1	0.948	6	0.952	11	0.950	16	0.947	21	0.953
2	0.956	7	0.941	12	0.957	17	0.959	22	0.955
3	0.936	8	0.943	13	0.943	18	0.946	23	0.937
4	0.939	9	0.964	14	0.945	19	0.957	24	0.963
5	0.946	10	0.952	15	0.945	20	0.952	25	0.946

then we can find

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i = \frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i = 0.949$$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]^2 = \frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i - 0.949\right]^2 = 5.788\times10^{-5}$

then $S = 7.608\times10^{-3}$

now, we can find from the standard normal distn.'s table two no's $\pm z_{1-\frac{\alpha}{2}}$,

s.t: $\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$

and the 95% C.I for the integral I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}}z_{0.95}\ ,\ \bar{\hat{\theta}} + \frac{S}{\sqrt{n}}z_{0.95}\right)$

$$\bar{\hat{\theta}} - \frac{S}{\sqrt{n}}z_{0.95} = 0.949 - \frac{7.608\times10^{-3}}{5}\cdot 1.645 = 0.947$$

$$\bar{\hat{\theta}} + \frac{S}{\sqrt{n}}z_{0.95} = 0.949 + \frac{7.608\times10^{-3}}{5}\cdot 1.645 = 0.952$$

Therefore the 95% C.I for I is: $(0.947\ ,\ 0.952)$.

Example (2.2.2.3): Calculation the 99.5 % C.I for the integral

$$I = \int_0^1\int_0^1\int_1^2\int_0^2\int_1^3 e^{x_1^2+x_2^2} + 2x_3 - \sin(x_4+x_5)\,dx_1dx_2dx_3dx_4dx_5\,.$$

Solution:

For the best no. of trails N with $p = 0.99,\ \varepsilon = 0.01$:

$$N \geq \frac{p(1-p)\left[c\,(b_1-a_1)(b_2-a_2)...(b_5-a_5)\right]}{(1-\alpha)\varepsilon^2} = 1\times10^4$$

By calculating $\hat{\theta}_2$, according to the Sample Mean Monte Carlo method with number of repetition n=25, and the result are tabulated in table(2.6):

Table(2.3)

"The Estimators of

$$I=\int_0^1\int_0^1\int_1^2\int_0^2\int_1^3 e^{x_1^2+x_2^2}+2x_3-\sin(x_4+x_5)\,dx_1dx_2dx_3dx_4dx_5\text{, using}$$

Sample Mean method"

n	$\hat{\theta}_2$	**n**	$\hat{\theta}_2$	**n**	$\hat{\theta}_2$	**n**	$\hat{\theta}_2$	**n**	$\hat{\theta}_2$
1	20.243	**6**	20.172	**11**	20.167	**16**	20.226	**21**	20.242
2	20.090	**7**	20.125	**12**	20.257	**17**	20.170	**22**	20.175
3	20.218	**8**	20.190	**13**	20.114	**18**	20.099	**23**	20.108
4	20.117	**9**	20.200	**14**	20.254	**19**	20.194	**24**	20.167
5	20.125	**10**	20.157	**15**	20.186	**20**	20.186	**25**	20.180

then we can find

$$\bar{\hat{\theta}}=\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i=\frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i=20.174$$

and $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i-\bar{\hat{\theta}}\right]^2=\frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i-20.174\right]^2=2.419\times10^{-3}$

now, we can find from the standard normal distn.'s table two no.s $\pm z_{1-\frac{\alpha}{2}}$,

s.t: $\Pr\left[-z_{1-\frac{\alpha}{2}}<Z<z_{1-\frac{\alpha}{2}}\right]=1-\alpha$

and the 99.5% C.I for the integral I is $\left(\bar{\hat{\theta}}-\frac{S}{\sqrt{n}}z_{0.995}\ ,\ \bar{\hat{\theta}}+\frac{S}{\sqrt{n}}z_{0.995}\right)$

$$\bar{\hat{\theta}}-\frac{S}{\sqrt{n}}z_{0.995}=20.174-\frac{0.049}{5}\cdot 2.6=20.149$$

$$\bar{\hat{\theta}}+\frac{S}{\sqrt{n}}z_{0.995}=20.174+\frac{0.049}{5}\cdot 2.6=20.200$$

Therefore the 99.5% C.I for I is: $(20.149\ ,\ 20.200)$

2.2.3 Efficiency of Monte Carlo Methods:

Suppose two Monte Carlo methods exist for estimating the integral I. let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimates produced by these methods s.t :

$$E\left[\hat{\theta}_1\right] = E\left[\hat{\theta}_2\right] = I \ .$$

We denote by t_1 and t_2 the units of computing time required for evaluating the r.v^s $\hat{\theta}_1$ and $\hat{\theta}_2$ respectively.

Let the variance associated with the 1st method be $\mathrm{var}(\hat{\theta}_1)$ and the associated with the 2nd method be $\mathrm{var}(\hat{\theta}_2)$, then we say that the 1st method is more efficient than the 2nd method if $eff. = \frac{t_1 \,\mathrm{var}(\hat{\theta}_1)}{t_2 \,\mathrm{var}(\hat{\theta}_2)} < 1.$

Let us compare now the efficiency of the hit or miss method with that of the sample mean method.

Proposition (2.2.3.1): $\mathrm{var}(\hat{\theta}_2) \le \mathrm{var}(\hat{\theta}_1)$

Proof:

Subtracting eq.(2.18) from eq.(2.7), we obtain:

$$\mathrm{var}(\hat{\theta}_1) - \mathrm{var}(\hat{\theta}_2) = \frac{1}{N}(b_1 - a_1)(b_2 - a_2)...(b_n - a_n)\left[cI - \int_{a_1}^{b_1}\int_{a_2}^{b_2}...\int_{a_n}^{b_n} g^2(x_1, x_2, ..., x_n)dx_1 dx_2 .. dx_n \right]$$

...(2.19)

Note that $g(x_1, x_2, ..., x_n) \le c$,

Therefore

$$cI - \int_{a_1}^{b_1}\int_{a_2}^{b_2}...\int_{a_n}^{b_n} g^2(x_1, x_2, ..., x_n)dx_1 dx_2 .. dx_n \ge 0$$

and further

$\mathrm{var}(\hat{\theta}_1) - \mathrm{var}(\hat{\theta}_2) \geq 0$

Assuming that the computing times t_1 and t_2 for $\hat{\theta}_1$ and $\hat{\theta}_2$ are approximately equal, we conclude that the sample mean method is more efficient than the hit or miss method.

If $\mathrm{var}(\hat{\theta}_1)$ and $\mathrm{var}(\hat{\theta}_2)$ are unknown, we can replace them by their estimators:

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} g(X_{1i}, X_{2i}, ..., X_{ni})(b_1 - a_1)(b_2 - a_2)...(b_n - a_n) - \theta\right]^2 \quad \text{…(2.20)}$$

and then estimated by

$$eff. = \frac{t_1 S_1^2}{t_2 S_2^2} \quad \text{…(2.21)}$$

By comparing examples (2.2.1.1) and (2.2.2.1), we can show that:

$$eff. = \frac{t_1 S_1^2}{t_2 S_2^2} = \frac{t_1(2.343 \times 10^{-4})}{t_2(4.277 \times 10^{-6})},$$

and by taking $t_1 \approx t_2$ then we get, $eff. = 547.814 > 1$.

Compare again examples (2.2.1.2) and (2.2.2.2), it is easily to evaluate

$$eff. = \frac{t_1 S_1^2}{t_2 S_2^2} = \frac{t_1(2.95 \times 10^{-4})}{t_2(5.788 \times 10^{-5})},$$

and take $t_1 \approx t_2$ then, $eff. = 5.097 > 1$.

Compare again examples (2.2.1.3) and (2.2.2.3), it is easily to evaluate

$$eff. = \frac{t_1 S_1^2}{t_2 S_2^2} = \frac{t_1(42.5 \times 10^{-3})}{t_2(2.419 \times 10^{-3})},$$

and take $t_1 \approx t_2$ then, $eff. = 17.569 > 1$.

which means that the sample mean method is more accurate than the hit or miss method.

It is interesting to note that, estimating the integral by $\hat{\theta}_1$ and $\hat{\theta}_2$, we do not need to know the function $g(x_1,x_2,\ldots,x_n)$ explicitly, we need only evaluate $g(x_1,x_2,\ldots,x_n)$ at any point $(x_1,x_2,\ldots,x_n)$.

CHAPTER 3

VARIANCE REDUCTION TECHNIQUS

3.1 Introduction:

In this chapter we shall discuss the variance reduction technique for estimating the n-dimensional integrals where four procedures of sampling are employed, namely, the Importance sampling, the Correlated sampling, Partition of the Region, and the Biased estimators.

Some related theorems, corollaries, and propositions to these procedures are proved and discuss in details.

3.2 Variance Reduction Techniques [15]:

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. One way to gain this information is through a direct crude simulation of the process. Results from this simulation can then be used to define variance reduction techniques that will refine and improve the efficiency of a 2nd simulation. Therefore the more that is known about the problem, the more effective the variance reduction techniques that can be employed. Hence it is always important to clearly define what is known about the problem. Knowledge of a process to be simulated can be qualitative, quantitative, or both.

3.2.1 Importance Sampling:

Let us consider the problem of estimating the n-dimensional integral

$$I=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\ldots\int_{a_n}^{b_n} g(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n \qquad , x\in\Omega\subset R^n \quad \ldots(3.1)$$

where $\Omega=\{(x_1,x_2,\ldots,x_n):a_i\le x_i\le b_i\ ,i=1,2,\ldots,n\}$

We suppose that $g\in L^2(x)$ ”in other words, that

$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\ldots\int_{a_n}^{b_n} g^2(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n$ exists and therefore that *I* exists”.

The basic idea of this technique [11] consists of concentrating the distn. of the sample point in the parts of the region Ω that are of most “importance” instead of spreading them out evenly. By analogy with eq.(2.14) and eq.(2.15) of chapter two we can represent the integral (3.1) as

$$I=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{g(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n)}f(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n$$

$$=E\left[\frac{g(X_1,X_2,\ldots,X_n)}{f(X_1,X_2,\ldots,X_n)}\right] \qquad \ldots(3.2)$$

Here $(X_1,X_2,\ldots,X_n)$ is any random vector with p.d.f $f(x_1,x_2,\ldots,x_n)$, s.t $f(x_1,x_2,\ldots,x_n)>0$, for each $(x_1,x_2,\ldots,x_n)\in\Omega$.

The function $f(x_1,x_2,\ldots,x_n)$ is called the importance sampling distn.. It is obvious from eq.(3.2) that if

$$\zeta=\frac{g(X_1,X_2,\ldots,X_n)}{f(X_1,X_2,\ldots,X_n)}$$

is an unbiased estimator for I, with the variance

$$\mathrm{var}(\zeta)=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n)}f(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n-I^2$$

$\ldots(3.3)$

In order to estimate the integral we take a samples $(X_{11},X_{21},\ldots,X_{n1})$, $(X_{12},X_{22},\ldots,X_{n2})$, …, $(X_{1N},X_{2N},\ldots,X_{nN})$ from the p.d.f $f(x_1,x_2,\ldots,x_n)$ and substitute its values in the sample-mean formula

$$\theta_3=\frac{1}{N}\sum_{i=1}^{N}\frac{g(X_{1i},X_{2i},\ldots,X_{ni})}{f(X_{1i},X_{2i},\ldots,X_{ni})} \qquad \ldots(3.4)$$

The necessary steps for estimating the integrals by the importance sampling technique can be describe by IS-Algorithm:

IS-Algorithm:

1. Generate a seq. $\{X_i\}_{i=1}^{n\cdot N}$ of $n.N$ random numbers which distributed with the p.d.f $f(x_1,x_2,\ldots,x_n)$.

2. Arrange the random numbers into N pairs $(X_{11},X_{21},\ldots,X_{n1})$, $(X_{12},X_{22},\ldots,X_{n2})$, …, $(X_{1N},X_{2N},\ldots,X_{nN})$ in any fashion s.t each random number X is used exactly once.
3. Estimate the integral by:

$$\hat{\theta}_3 = \frac{1}{N}\sum_{i=1}^{N}\frac{g(X_{1i},X_{2i},\ldots,X_{ni})}{f(X_{1i},X_{2i},\ldots,X_{ni})}.$$

For illustration we shall solve the following example

Example (3.2.1.1): Calculating the 99.5% C.I for the integral

$$I = \int_0^1 e^{-x^2}\,dx \quad .$$

Solution:

For the best no. of trials N with $p = 0.99,\ \varepsilon = 0.001$:

$$N \geq \frac{p(1-p)\left[c(b-a)\right]}{(1-\alpha)\varepsilon^2} = 4\times 10^4$$

By calculating $\hat{\theta}_3$ according to the Importance Sampling technique with using the standard normal distn. as an importance sampling distn., and the number of repetition n=25, and the result are tabulated in table(3.1):

Table(3.1)

"The Estimators of $I = \int_0^1 e^{-x^2} dx$, using The Importance Sampling Technique"

n	$\hat{\theta}_3$	n	$\hat{\theta}_3$	n	$\hat{\theta}_3$	n	$\hat{\theta}_3$	n	$\hat{\theta}_3$
1	0.737	6	0.737	11	0.736	16	0.739	21	0.737
2	0.737	7	0.739	12	0.736	17	0.737	22	0.738
3	0.736	8	0.737	13	0.737	18	0.738	23	0.739
4	0.737	9	0.734	14	0.733	19	0.737	24	0.736
5	0.735	10	0.737	15	0.737	20	0.737	25	0.736

then we can find

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i = \frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i = 0.73676$$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]^2 = \frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i - 0.746\right]^2 = 1.94\times10^{-6}$

then $S = 1.393\times10^{-3}$

now, we can find from the standard normal distn.'s table two no.s $\pm z_{1-\frac{\alpha}{2}}$,

s.t: $\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$

and the 99.5% C.I for the integral I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} \,,\, \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995}\right)$

$$\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} = 0.73676 - \frac{1.393\times10^{-3}}{5}\cdot 2.6 = 0.736036$$

$$\bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995} = 0.73676 + \frac{1.393\times10^{-3}}{5}\cdot 2.6 = 0.737484$$

Therefore the 99.5% C.I for I is: $(0.736036 \,,\, 0.737484)$

We now show how to choose the distn. of the random vector $(X_1,X_2,\dots,X_n)$ in order to minimize the variance of ζ, which is the same as to minimize the variance of θ_3.

Theorem (3.1.1):

The minimum of var(ζ) is equal to

$$\operatorname{var}(\zeta_0)=\left[\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\left|g(x_1,x_2,\dots,x_n)\right|dx_1dx_2\dots dx_n\right]^2-I^2 \qquad \dots(3.5)$$

and occurs when the random vector $(X_1,X_2,\dots,X_n)$ is distributed with p.d.f

$$f(x_1,x_2,\dots,x_n)=\frac{\left|g(x_1,x_2,\dots,x_n)\right|}{\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\left|g(x_1,x_2,\dots,x_n)\right|dx_1dx_2\dots dx_n} \qquad \dots(3.6)$$

Proof:

The formula of eq.(3.5) follows directly if we substitute eq.(3.6) into eq.(3.3).

In order to prove that $\operatorname{var}(\zeta_0)\le\operatorname{var}(\zeta)$, it is enough to prove that

$$\left[\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\left|g(x_1,x_2,\dots,x_n)\right|dx_1dx_2\dots dx_n\right]^2\le\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{g^2(x_1,x_2,\dots,x_n)}{f(x_1,x_2,\dots,x_n)}dx_1dx_2\dots dx_n \qquad \dots(3.7)$$

but

$$\left[\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\left|g(x_1,x_2,\ldots,x_n)\right|dx_1dx_2\ldots dx_n\right]^2 =$$

$$\left[\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{\left|g(x_1,x_2,\ldots,x_n)\right|}{\left[f(x_1,x_2,\ldots,x_n)\right]^{1/2}}\left[f(x_1,x_2,\ldots,x_n)\right]^{1/2}dx_1dx_2\ldots dx_n\right]^2$$

and by extended Cauchy-Schwarz inequality

$$\left[\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{\left|g(x_1,x_2,\ldots,x_n)\right|}{\left[f(x_1,x_2,\ldots,x_n)\right]^{1/2}}\left[f(x_1,x_2,\ldots,x_n)\right]^{1/2}dx_1dx_2\ldots dx_n\right]^2 \le$$

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n)}dx_1dx_2\ldots dx_n\cdot\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}f(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n$$

$$=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n)}dx_1dx_2\ldots dx_n \qquad \ldots(3.8)$$

Corollary (3.1.1):

If $g(x_1,x_2,\ldots,x_n)>0$, then the optimal p.d.f is

$$f(x_1,x_2,\ldots,x_n)=\frac{g(x_1,x_2,\ldots,x_n)}{I} \qquad \ldots(3.9)$$

and var(ζ) =0.

It has been shown that [15] that this method is unfortunately useless, since the optimal density contains the integral $\int\int\cdots\int\left|g(x_1,x_2,\ldots,x_n)\right|dx_1dx_2\ldots dx_n$, which is practically equivalent to computing *I*. In the case where $g(x_1,x_2,\ldots,x_n)$ has a constant sign it is precisely equivalent to calculating *I*. But if we already have *I*, we do not need Monte Carlo methods to estimate it.

In particular, if we choose $f(x_1,x_2,\ldots,x_n)$ as a uniform density function with

$$f(x_i)=\frac{1}{b_i-a_i},\ a_i \le x_i \le b_i$$
$$=0,\ e.w$$

then we will get the Monte Carlo sample mean method itself.

Not all is lost, however. The variance can be essentially reduced if $f(x_1,x_2,\ldots,x_n)$ is chosen in order to have a shape similar to that of $|g(x_1,x_2,\ldots,x_n)|$. When choosing $f(x_1,x_2,\ldots,x_n)$ in such a way we have to take into consideration the difficulties of sampling from such a p.d.f, especially if $|g(x_1,x_2,\ldots,x_n)|$ is not a well behaved function.

Consider the problem of choosing the parameters of the distn. $f(x_1,x_2,\ldots,x_n)$ in an optimal way. We assume that the p.d.f $f(x_1,x_2,\ldots,x_n)$ is determined up to the vector of parameters α, that is $f(x_1,x_2,\ldots,x_n)=f(x_1,x_2,\ldots,x_n,\alpha)$. For instance, if $f(x_1,x_2)$ represents two-dimensional Normal distn., that is $(X_1,X_2)\sim N(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$, then the unknown parameters can be the values of μ_1, μ_2, σ_1, σ_2 and ρ.

Generally if we want to choose the vector of parameters α to minimize the variance of θ_3, that is

$$\min_{\alpha} \operatorname{var}\left[\theta_3=\frac{1}{N}\sum_{i=1}^{N}\frac{g(x_{1i},x_{2i},\ldots,x_{ni})}{f(x_{1i},x_{2i},\ldots,x_{ni},\alpha)}\right]=$$
$$\frac{1}{N}\min_{\alpha}\left[\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n,\alpha)}dx_1dx_2\ldots dx_n - I^2\right]$$

…(3.10)

Which equivalent to

$$\min_{\alpha}\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n,\alpha)}dx_1dx_2\ldots dx_n \quad \ldots(3.11)$$

and the function

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}\frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n,\alpha)}dx_1dx_2\ldots dx_n \qquad \ldots(3.12)$$

can be multiextremal and generally it is difficult to find the optimal α.

3.2.2 Correlated Sampling [15]:

Correlated sampling is one of the most powerful variance reduction techniques.

Frequently, the primary objective of a simulation study is to determine the effect of a small change in the system. The sample mean Monte Carlo method would make two independent runs, with and without the change in the system being simulated, and subtract the results obtained. Unfortunately, the difference being calculated is often small compared to the separate results, while the variance of the difference will be the sum of the variance in the two runs, which is usually significant. If, instead of being independent, the two simulations use the same random numbers, the results can be highly positively correlated, which provides a reduction in the variance. Another way of viewing correlated sampling through random numbers control is to realize that the use of the same random numbers generates identically histories in those parts of the two system, that are the same. Thus the aim of the correlated sampling is to produce a high positive correlated between two similar processes so that the variance of the difference is considerably smaller than it would be if the two processes were statistically independent.

Unfortunately, there is no general procedure that can be implemented in correlated sampling. However, in the following two situations correlated sampling can be successfully employed:

1. The value of the small change in a system is to be calculated.

2. The difference in a parameter in two or more similar cases is of more interest than its absolute value.

Let us assume that we desire to estimate

$\Delta I = I_1 - I_2$

where

$$I_1 = \int\int ... \int g_1(x_1, x_2, ..., x_n)\, f_1(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n \qquad ...(3.13)$$

$$I_2 = \int\int ... \int g_2(x_1, x_2, ..., x_n)\, f_2(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n \qquad ...(3.14)$$

Then the procedure for correlated sampling is as follows:

1. Generate $(X_{11}, X_{21}, \ldots, X_{n1}), \ldots, (X_{1N}, X_{2N}, \ldots, X_{nN})$ from $f_1(x_1, x_2, \ldots, x_n)$ and $(Y_{11}, Y_{21}, \ldots, Y_{n1}), \ldots, (Y_{1N}, Y_{2N}, \ldots, Y_{nN})$ from $f_2(x_1, x_2, \ldots, x_n)$.

2. Estimate ΔI using

$$\Delta\theta = \frac{1}{N}\sum_{i=1}^{N} g_1(X_{1i}, X_{2i}, ..., X_{ni}) - \frac{1}{N}\sum_{i=1}^{N} g_2(Y_{1i}, Y_{2i}, ..., Y_{ni})$$

$$= \frac{1}{N}\sum_{i=1}^{N} \Delta_i , \qquad ...(3.15)$$

where $\Delta_i = g_1(X_{1i}, X_{2i}, ..., X_{ni}) - g_2(X_{1i}, X_{2i}, ..., X_{ni}) \quad , \forall i = 1, 2, ..., N$

the variance of $\Delta\theta$ is

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\,\mathrm{cov}(\hat{\theta}_1, \hat{\theta}_2) \qquad ...(3.16)$$

where

$$\hat{\theta}_1 = \frac{1}{N}\sum_{i=1}^{N} g_1(X_{1i}, X_{2i}, ..., X_{ni}),$$

$$\hat{\theta}_2 = \frac{1}{N}\sum_{i=1}^{N} g_2(Y_{1i}, Y_{2i}, ..., Y_{ni}),$$

$$\sigma_1^2 = E\left[\hat{\theta}_1 - I_1\right]^2,$$

$$\sigma_2^2 = E\left[\hat{\theta}_2 - I_2\right]^2, \text{ and}$$

$$\mathrm{cov}(\hat{\theta}_1,\hat{\theta}_2)=E\left[(\hat{\theta}_1-I_1)(\hat{\theta}_2-I_2)\right].$$

Now, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are statistically independent, then $\mathrm{cov}(\hat{\theta}_1,\hat{\theta}_2)=0$, and $\sigma^2=\sigma_1^2+\sigma_2^2$.

However, if the random vectors X and Y are positively correlated and if $g_1(x_1,x_2,\dots,x_n)$ is similar to $g_2(x_1,x_2,\dots,x_n)$ in shape, then the r.v^s $\hat{\theta}_1$ and $\hat{\theta}_2$ will also be positively correlated, that is, $\mathrm{cov}(\hat{\theta}_1,\hat{\theta}_2)>0$, and the variance of $\Delta\theta$ may be greatly reduced.

Thus the key to reducing the variance of $\Delta\theta$ is to insure positive correlation between the estimates $\hat{I}_1$ and $\hat{I}_2$. This can be achieved in several ways. The easiest way is to obtain correlated samples through random number control. Specifically, this can be accomplished by using the same "common" sequence of random vectors $(U_{11},U_{21},\dots,U_{n1})$, $(U_{12},U_{22},\dots,U_{n2})\dots$, $(U_{1N},U_{2N},\dots,U_{nN})$ in both simulations, that is, the sequences $(X_{11},X_{21},\dots,X_{n1})$, $(X_{12},X_{22},\dots,X_{n2}),\dots$, $(X_{1N},X_{2N},\dots,X_{nN})$ and $(Y_{11},Y_{21},\dots,Y_{n1})$, $(Y_{12},Y_{22},\dots,Y_{n2}),\dots$, $(Y_{1N},Y_{2N},\dots,Y_{nN})$ are generated using $X_i=F_1^{-1}(U_{1i},U_{2i},\dots,U_{ni})$ and $Y_i=F_2^{-1}(U_{1i},U_{2i},\dots,U_{ni})$ respectively. Clearly, if f_1 is similar to f_2, the r.v^s X_i and Y_i will be highly positively correlated since they both used the same random numbers.

It has been note that [15] it is difficult to be specific as to how random number control should be applied generally. As a rule, however, to achieve maximum correlation common random number^s should be used whenever the similarities in problem structure will permit this.

3.2.3 Partition of the Region [15]:

In this technique [17] we break the region

$$\Omega=\{(x_1,x_2,\dots,x_n):a_i<x_i<b_i\ ,i=1,2,\dots,n\},$$

into two parts $\Omega = \Omega_1 \bigcup \Omega_2$,where

$\Omega_1 = \{(x_1, x_2, ..., x_n) : a_i \le x_i \le c_i \ , i = 1, 2, ..., n\}$, and

$\Omega_2 = \{(x_1, x_2, ..., x_n) : c_i \le x_i \le b_i \ , i = 1, 2, ..., n\}$,

Representing the integral I as

$$I = \int_{a_1}^{b_1}\int_{a_2}^{b_2}...\int_{a_n}^{b_n} g(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n$$

$$= \int_{a_1}^{c_1}\int_{a_2}^{c_2}...\int_{a_n}^{c_n} g(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n$$

$$+ \int_{c_1}^{b_1}\int_{c_2}^{b_2}...\int_{c_n}^{b_n} g(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n \qquad ...(3.17)$$

Let us assume the integral

$$I_1 = \int_{a_1}^{c_1}\int_{a_2}^{c_2}...\int_{a_n}^{c_n} g(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n \qquad ...(3.18)$$

can be calculated analytically, and let us define a truncated p.d.f

$$h(x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{1-p} \quad , if \ (x_1, x_2, \ldots, x_n) \in \Omega_2$$
$$= 0 \quad , e.w \qquad ...(3.19)$$

where $p = \int_{a_1}^{c_1}\int_{a_2}^{c_2}...\int_{a_n}^{c_n} f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n$.

Formula (3.17) can be written as

$$I = I_1 + \int_{c_1}^{b_1}\int_{c_2}^{b_2}...\int_{c_n}^{b_n} g(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n$$

$$= I_1 + \int_{a_1}^{c_1}\int_{a_2}^{c_2}...\int_{a_n}^{c_n} \frac{g(x_1, x_2, ..., x_n)}{h(x_1, x_2, ..., x_n)} h(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n$$

$$= I_1 + E\left[\frac{g(X_1, X_2, ..., X_n)}{h(X_1, X_2, ..., X_n)}\right]$$

$$= I_1 + (1-p)E\left[\frac{g(X_1, X_2, ..., X_n)}{f(X_1, X_2, ..., X_n)}\right] \qquad ...(3.20)$$

an unbiased estimator for I is then

$$Y = I_1 + (1-p)\frac{g(X_1, X_2, \ldots, X_n)}{f(X_1, X_2, \ldots, X_n)} \qquad \ldots(3.21)$$

and the integral I can be estimated by

$$\hat{\theta}_4 = I_1 + (1-p)\frac{1}{N}\sum_{i=1}^{N}\frac{g(X_1, X_2, \ldots, X_n)}{f(X_1, X_2, \ldots, X_n)} \qquad \ldots(3.22)$$

The necessary steps for estimating the integrals by the partition of the region technique can be describe by PR-Algorithm:

PR-Algorithm:

1. Read $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n, c_1, c_2, \ldots,$ and c_n.
2. Compute:

   $I_1 = \int_{a_1}^{c_1}\int_{a_2}^{c_2}\ldots\int_{a_n}^{c_n} g(x_1, x_2, \ldots, x_n)dx_1 dx_2 \ldots dx_n$, and

   $p = \int_{a_1}^{c_1}\int_{a_2}^{c_2}\ldots\int_{a_n}^{c_n} g(x_1, x_2, \ldots, x_n)dx_1 dx_2 \ldots dx_n$, where $f(x_1, x_2, \ldots, x_n)$ is a p.d.f.
3. Generate a seq. $\{X_i\}_{i=1}^{nN}$ of nN random numbers which distributed with the p.d.f $h(x_1, x_2, \ldots, x_n) = \dfrac{f(x_1, x_2, \ldots, x_n)}{1-p}$.
4. Arrange the random numbers into N pairs $(X_{11}, X_{21}, \ldots, X_{n1})$, $(X_{12}, X_{22}, \ldots, X_{n2})$, …, $(X_{1N}, X_{2N}, \ldots, X_{nN})$ in any fashion s.t each random number X is used exactly once.
5. Estimate the integral by:

$$\hat{\theta}_4 = I_1 + (1-p)\frac{1}{N}\sum_{i=1}^{N}\frac{g(X_{1i}, X_{2i}, \ldots, X_{ni})}{f(X_{1i}, X_{2i}, \ldots, X_{ni})}.$$

For illustration we will solve the following example

Example (3.2.3.1): Calculating the 99.5% C.I for the integral

$$I = \int_0^1 e^{-x^2} dx = \int_0^{1/2} e^{-x^2} dx + \int_{1/2}^1 e^{-x^2} dx = I_1 + I_2 \quad .$$

Solution:

For the best no. of trails N with $p = 0.99, \varepsilon = 0.001$:

$$N \geq \frac{p(1-p)\left[c(b-a)\right]}{(1-\alpha)\varepsilon^2} = 4\times 10^4$$

By calculating $\hat{\theta}_4$ according the Partition of the Region technique, and using the standard normal distn. as an importance sampling distn., with I_1=0.461 which calculated by MathCAD standard forms, and the number of repetition $n=25$, and the results are tabulated in table(3.2):

Table(3.2)

"The Estimators of $I = \int_0^1 e^{-x^2} dx$, using The Partition of The Region Technique"

n	$\hat{\theta}_4$	n	$\hat{\theta}_4$	n	$\hat{\theta}_4$	n	$\hat{\theta}_4$	n	$\hat{\theta}_4$
1	0.736	6	0.735	11	0.738	16	0.737	21	0.736
2	0.736	7	0.737	12	0.737	17	0.739	22	0.738
3	0.737	8	0.738	13	0.734	18	0.737	23	0.737
4	0.737	9	0.739	14	0.736	19	0.740	24	0.737
5	0.739	10	0.737	15	0.737	20	0.737	25	0.739

then we can find

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i = \frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i = 0.7372$$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]^2 = \frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i - 0.746\right]^2 = 1.83\times10^{-6}$

then $S = 1.354\times10^{-3}$

now, we can find from the standard normal distn.'s table two no.s $\pm z_{1-\frac{\alpha}{2}}$,

s.t: $\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$

and the 99.5% C.I for the integral I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} \,,\, \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995}\right)$

$$\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} = 0.7372 - \frac{1.354\times10^{-3}}{5}\cdot 2.6 = 0.736496$$

$$\bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995} = 0.7372 + \frac{1.354\times10^{-3}}{5}\cdot 2.6 = 0.737904$$

Therefore the 99.5% C.I for I is: $(0.736496 \,,\, 0.737904)$

Proposition(3.2.3.1): $\operatorname{var}(\hat{\theta}_4) \le (1-p)\operatorname{var}(\hat{\theta}_3)$

Proof:

We have from eq.(3.4) that

$$N \operatorname{var}(\hat{\theta}_3) = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} \frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n)}\,dx_1 dx_2 \ldots dx_n - I^2$$

$$= \int_{a_1}^{c_1}\int_{a_2}^{c_2}\cdots\int_{a_n}^{c_n} \frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n)}\,dx_1 dx_2 \ldots dx_n$$

$$+ \int_{c_1}^{b_1}\int_{c_2}^{b_2}\cdots\int_{c_n}^{b_n} \frac{g^2(x_1,x_2,\ldots,x_n)}{f(x_1,x_2,\ldots,x_n)}\,dx_1 dx_2 \ldots dx_n - I^2 \qquad \ldots(3.23)$$

and correspondingly, from eq.(3.22) that

$$N\ \mathrm{var}(\hat{\theta}_4) = (1-p)^2 \int_{c_1}^{b_1}\int_{c_2}^{b_2}\cdots\int_{c_n}^{b_n} \frac{g^2(x_1,x_2,\dots,x_n)}{f^2(x_1,x_2,\dots,x_n)}\,\frac{f(x_1,x_2,\dots,x_n)}{(1-p)}dx_1dx_2\dots dx_n$$
$$-\left[(1-p)\int_{c_1}^{b_1}\int_{c_2}^{b_2}\cdots\int_{c_n}^{b_n} \frac{g(x_1,x_2,\dots,x_n)}{f(x_1,x_2,\dots,x_n)}\,\frac{f(x_1,x_2,\dots,x_n)}{(1-p)}dx_1dx_2\dots dx_n\right]^2$$

$$= (1-p)\int_{c_1}^{b_1}\int_{c_2}^{b_2}\cdots\int_{c_n}^{b_n} \frac{g^2(x_1,x_2,\dots,x_n)}{f(x_1,x_2,\dots,x_n)}\,dx_1dx_2\dots dx_n$$
$$-\left[\int_{c_1}^{b_1}\int_{c_2}^{b_2}\cdots\int_{c_n}^{b_n} g(x_1,x_2,\dots,x_n)dx_1dx_2\dots dx_n\right]^2 \qquad \dots(3.25)$$

Multiplying eq.(3.23) by (1-p) and subtracting eq.(3.25), we obtain

$$N\left[(1-p)\,\mathrm{var}(\hat{\theta}_3) - \mathrm{var}(\hat{\theta}_4)\right] = (1-p)\int_{a_1}^{c_1}\int_{a_2}^{c_2}\cdots\int_{a_n}^{c_n} \frac{g^2(x_1,x_2,\dots,x_n)}{f(x_1,x_2,\dots,x_n)}\,dx_1dx_2\dots dx_n$$
$$-(1-p)I^2 + \left[\int_{c_1}^{b_1}\int_{c_2}^{b_2}\cdots\int_{c_n}^{b_n} g(x_1,x_2,\dots,x_n)dx_1dx_2\dots dx_n\right]^2$$

Therefore

$$N\left[(1-p)\,\mathrm{var}(\hat{\theta}_3) - \mathrm{var}(\hat{\theta}_4)\right] = (1-p)\int_{a_1}^{c_1}\int_{a_2}^{c_2}\cdots\int_{a_n}^{c_n} \frac{g^2(x_1,x_2,\dots,x_n)}{f(x_1,x_2,\dots,x_n)}\,dx_1dx_2\dots dx_n$$
$$-(1-p)I^2 + (I-I_1)^2$$

$$\dots(3.26)$$

Now, introducing

$$c^2 = \int_{a_1}^{c_1}\int_{a_2}^{c_2}\cdots\int_{a_n}^{c_n} \frac{g^2(x_1,x_2,\dots,x_n)}{f(x_1,x_2,\dots,x_n)}\,dx_1dx_2\dots dx_n - \frac{I_1^2}{p}$$
$$= \int_{a_1}^{c_1}\int_{a_2}^{c_2}\cdots\int_{a_n}^{c_n} \left(\frac{g(x_1,x_2,\dots,x_n)}{f(x_1,x_2,\dots,x_n)} - \frac{I_1}{p}\right)^2 f(x_1,x_2,\dots,x_n)\,dx_1dx_2\dots dx_n$$

$$\dots(3.27)$$

we have

$$N\left[(1-p)\operatorname{var}(\hat{\theta}_3)-\operatorname{var}(\hat{\theta}_4)\right]=(1-p)c^2+\left(p^{1/2}I-p^{-1/2}I_1\right)^2\geq 0$$

and the proposition is proved.

As a result of a proposition, we find that this technique is at least $(1\text{-}p)^{-1}$ times more efficient than the sample mean Monte Carlo method.

Practically, from examples (3.2.1.1) and (3.2.3.1):

$\operatorname{var}(\hat{\theta}_3)\approx S^2=1.94\times10^{-6}$, $\operatorname{var}(\hat{\theta}_4)\approx S^2=1.83\times10^{-6}$, and $p=0.99$

then $(1-p)\operatorname{var}(\hat{\theta}_3)=(0.01)\cdot(1.94\times10^{-6})1.94\times10^{-8}\leq 1.83=\operatorname{var}(\hat{\theta}_4)$

3.2.4 Biased Estimators:

Until now we have considered unbiased estimators for computing integrals. Using biased estimators, we can some times achieve useful results.

Let us estimate the integral

$$I=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}g(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n \qquad \ldots(3.28)$$

by

$$\hat{\theta}_5=\frac{\sum_{i=1}^{N}g(U_{1i},U_{2i},\ldots,U_{ni})}{\sum_{i=1}^{N}f(U_{1i},U_{2i},\ldots,U_{ni})} \qquad \ldots(3.29)$$

Instead of using the usual sample mean estimator

$$\hat{\theta}_3=\frac{1}{N}\sum_{i=1}^{N}\frac{g(X_{1i},X_{2i},\ldots,X_{ni})}{f(X_{1i},X_{2i},\ldots,X_{ni})},$$

Here U is distributed uniformly in Ω, that is

$$h(u)=\begin{cases}\dfrac{1}{V} & ,if\ u\in\Omega\\ =0 & ,e.w\end{cases}$$

where $V = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} dx_1 dx_2 \ldots dx_n$ …(3.30)

and $(X_1,X_2,\ldots,X_n)$ is distributed according to $f(x_1,x_2,\ldots,x_n)$.

The necessary steps for estimating the integrals using the biased estimator technique can be describe by BE-Algorithm:

BE-Algorithm:

1. Generate a seq. $\{U_i\}_{i=1}^{nN}$ of nN uniform random numbers.
2. Arrange the random numbers into N pairs $(U_{11},U_{21},\ldots,U_{n1})$, $(U_{12},U_{22},\ldots U_{n2})$, …, $(U_{1N},U_{2N},\ldots U_{nN})$ in any way s.t each random number U is used exactly once.
3. Estimate the integral by:

$$\hat{\theta}_5 = \frac{\sum_{i=1}^{N} g(U_{1i},U_{2i},\ldots,U_{ni})}{\sum_{i=1}^{N} f(U_{1i},U_{2i},\ldots,U_{ni})}.$$

and to show the reduction of this method we will solve the following example:

Example (3.2.4.1): Calculating the 99.5% C.I for the integral

$$I = \int_0^1 e^{-x^2} dx \quad .$$

Solution:

For the best no. of trails N with $p = 0.99,\ \varepsilon = 0.001$:

$$N \geq \frac{p(1-p)\left[c(b-a)\right]}{(1-\alpha)\varepsilon^2} = 4\times 10^4$$

By calculating $\hat{\theta}_5$ according to the Biased Estimator method, with using the standard normal distn. as an importance sampling distn., and the number of repetition *n=25*, and the result are tabulated in table (3.3):

Table(3.3)

"The Estimator of $I = \int_0^1 e^{-x^2}\,dx$, using The Biased Estimator Technique"

n	$\hat{\theta}_5$	n	$\hat{\theta}_5$	n	$\hat{\theta}_5$	n	$\hat{\theta}_5$	n	$\hat{\theta}_5$
1	0.737	**6**	0.739	**11**	0.738	**16**	0.734	**21**	0.737
2	0.737	**7**	0.735	**12**	0.736	**17**	0.737	**22**	0.740
3	0.74	**8**	0.737	**13**	0.741	**18**	0.735	**23**	0.735
4	0.736	**9**	0.738	**14**	0.739	**19**	0.737	**24**	0.734
5	0.741	**10**	0.731	**15**	0.737	**20**	0.734	**25**	0.737

then we can find

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i = \frac{1}{25}\sum_{i=1}^{25}\hat{\theta}_i = 0.737$$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]^2 = \frac{1}{24}\sum_{i=1}^{25}\left[\hat{\theta}_i - 0.746\right]^2 = 4.833\times10^{-6}$

then $S = 2.198\times10^{-3}$

now, we can find from the standard normal distn.'s table two no.s $\pm z_{1-\frac{\alpha}{2}}$,

s.t: $\Pr\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$

and the 99.5% C.I for the integral I is $\left(\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995}\ ,\ \bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995}\right)$

$$\bar{\hat{\theta}} - \frac{S}{\sqrt{n}} z_{0.995} = 0.737 - \frac{2.198\times10^{-3}}{5}\cdot 2.6 = 0.736$$

$$\bar{\hat{\theta}} + \frac{S}{\sqrt{n}} z_{0.995} = 0.737 + \frac{2.198\times 10^{-3}}{5} \cdot 2.6 = 0.738$$

Therefore the 99.5% C.I for I is: $(0.736\ ,\ 0.738)$

it is clear that $E\left[\hat{\theta}_5\right] \neq I$, that is $\hat{\theta}_5$ is a biased estimator for I. let us show that $\hat{\theta}_5$ is consistent. And to prove consistency let us represent $\hat{\theta}_5$ as a ratio of two r.v^s $\hat{\theta}_5'$ and $\hat{\theta}_5''$, that is

$$\hat{\theta}_5 = \frac{\hat{\theta}_5'}{\hat{\theta}_5''} = \frac{\left(V/N\right)\sum_{i=1}^{N} g(U_{1i}, U_{2i}, ..., U_{ni})}{\left(V/N\right)\sum_{i=1}^{N} f(U_{1i}, U_{2i}, ..., U_{ni})} \qquad \ldots(3.31)$$

where

$$\hat{\theta}_5' = \frac{V}{N}\sum_{i=1}^{N} g(U_{1i}, U_{2i}, ..., U_{ni}) \qquad \ldots(3.32)$$

and

$$\hat{\theta}_5'' = \frac{V}{N}\sum_{i=1}^{N} f(U_{1i}, U_{2i}, ..., U_{ni}) \qquad \ldots(3.33)$$

further

$$E\left[\hat{\theta}_5'\right] = V\int_{a_1}^{b_1}\int_{a_2}^{b_2}\ldots\int_{a_n}^{b_n} g(u_1, u_2, ..., u_n)\, h(u_1, u_2, ..., u_n)\, du_1 du_2 ... du_n$$

$$E\left[\hat{\theta}_5'\right] = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\ldots\int_{a_n}^{b_n} g(u_1, u_2, ..., u_n)\, du_1 du_2 ... du_n \qquad \ldots(3.34)$$

and

$$E\left[\hat{\theta}_5''\right] = V\int_{a_1}^{b_1}\int_{a_2}^{b_2}\ldots\int_{a_n}^{b_n} f(u_1, u_2, ..., u_n)\, h(u_1, u_2, ..., u_n)\, du_1 du_2 ... du_n$$

$$= \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} g(u_1,u_2,\ldots,u_n)\,du_1\,du_2\ldots du_n \qquad \ldots(3.35)$$

With these results in hand we conclude that $\hat{\theta}_5'$ and $\hat{\theta}_5''$ converge to I and 1, respectively, when $N \to \infty$, which also means that

$$\lim_{N\to\infty}\left[\frac{\sum_{i=1}^{N} g(U_{1i},U_{2i},\ldots,U_{ni})}{\sum_{i=1}^{N} f(U_{1i},U_{2i},\ldots,U_{ni})}\right] \xrightarrow{a.s} I$$

if , $\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} \left|g(x_1,x_2,\ldots,x_n)\right| dx_1\,dx_2\ldots dx_n < \infty$ …(3.36)

and this shows that $\hat{\theta}_5$ is a consistent estimator of I.

The bias of $\hat{\theta}_5$ follows from

$$E\left[\hat{\theta}_5\right] = E\left[\frac{\sum_{i=1}^{N} g(U_{1i},U_{2i},\ldots,U_{ni})}{\sum_{i=1}^{N} f(U_{1i},U_{2i},\ldots,U_{ni})}\right] \neq \frac{E\left[\sum_{i=1}^{N} g(U_{1i},U_{2i},\ldots,U_{ni})\right]}{E\left[\sum_{i=1}^{N} f(U_{1i},U_{2i},\ldots,U_{ni})\right]} = I$$

…(3.37)

One major advantage of this method is that the sample is taken from the uniform distn. rather that from a general $f(x_1,x_2,\ldots,x_n)$ from which the generation of r.v^s can be difficult "recall for instance that is importance sampling $f(x_1,x_2,\ldots,x_n)$ has to be proportional to $\left|g(x_1,x_2,\ldots,x_n)\right|$, and if $g(x_1,x_2,\ldots,x_n)$ is a complicated function, it is difficult to generate from $f(x_1,x_2,\ldots,x_n)$ ".

Powell and Swann [13] called this method weighted uniform sampling. They showed that for sufficiently large N this method is $N^{1/2}$ times more efficient than the sample mean method.

3.3 Mean Square Error Comparison:

Table(3.4) show the deference between the M.S.E and the confidence intervals of the Monte Carlo methods and the variance reduction techniques for the integral $I = \int_0^1 e^{-x^2} dx$.

Table(3.4)

M.S.E Comparison

	M.S.E	***C.I***	***length***
HM	2.343×10^{-5}	(0.743000 , 0.749000)	0.006000
SM	4.277×10^{-6}	(0.746000 , 0.748000)	0.002000
IS	1.94×10^{-6}	(0.736036 , 0.737484)	0.001448
PR	1.83×10^{-6}	(0.736496 , 0.737904)	0.001408
BE	2.45×10^{-4}	(0.736000 , 0.738000)	0.002000

Conclutions

1. The sample mean method is more efficient than the hit or miss method where the estimators for both methods are unbiased, but its shown theoretically and practically that the variance of the sample mean estimator less than the variance of the hit or miss estimator.
2. The advantage of the sample mean Monte Carlo method that it is needs N random variants, while the hit or miss Monte Carlo method method need to $2N$ random variants for estimating the integrals and that save time and less storage in the computer memory.
3. The disadvantage of the Monte Carlo methods are:
   i. Monte Carlo methods are depends completely on generating pseudorandom variates which might carry dirty data.
   ii. The accuracy of both methods decreases when the dimension of integrals increases.
4. The usage of variance reduction techniques lead to higher accuracy and small confidence intervals for estimating integrals.

Republic of Iraq
Ministry of Higher Education
and Scientific Research
Al-Nahrain University

MONTE CARLO INTEGRATIONS AND VARIANCE REDUCTION TECHNIQUES FOR N-DIMENSIONAL INTEGRALS

A Thesis

Submitted to the College of Science\ Al-Nahrain University as a Partial Fulfillment of the Requirements for the Degree of Master of Science in Applied Mathematics

By

Akram A. Al-Sabbagh

[B. Sc., Al-Nahrain University, 2005]

Supervised by

Dr. Akram M. Al-Abood

April
2008

Rabe'e AlAwal
1429

Future Work

1. The errors produced from the estimation of integrals by Monte Carlo methods or by variance reduction techniques are a r.v'^{s} and it must have a distn.. This distn. can be approximated to a well-known distn. by using statistical methods, such as, Chi square goodness of fit test, Kolmogorov-Smirnov goodness of fit test, Cramer-Von Miser goodness of fit test, …etc.
2. Finding new techniques for estimating integrals with noise which can be compared with the techniques given in this thesis.
3. Solving the difficulties of finding p.d.f which have the same shape with the integrand function in estimating integrals by variance reduction techniques.
4. Finding methods for evaluating improper integrals by Monte Carlo simulation.

Introduction:

The Monte Carlo method is a method for approximately solving mathematical and real life problems by simulation of random quantities.

Historically, Monte Carlo methods are considered as a technique, using random or pseudorandom no.^{s} to solve a certain models, these random no.^{s} are essentially independent r.v.^{s} uniformly distributed over the interval [0,1].

In the 2nd half of the 19th century (1873), one of the earliest problems connected with Monte Carlo method is the famous "Buffon's needle problem" where it was found that the probability of a needle of length L thrown randomly onto a floor composed of parallel planks of width $D>L$ is $P = 2L/\pi D$ which can be estimated as the ratio of the no. of throws hitting the crack to the total no. of throws. In the beginning of the 20th century, the Monte Carlo was used to examine the Boltzmann equation. In (1908) the famous statistician W. S. Gosset "student" used the Monte Carlo method "experimental sampling" for estimating the correlated coefficient in his t-distribution [٢٠].

The term "Monte Carlo" was introduced by Van Neumann and Ulam during the World War II (1944) as a secret code name for a secret work at Los-Alamos involving research related to the atomic bomb "H-bomb". The name comes from the city of Monte Carlo the capital of the principality of the Monaco, famous for it's gambling house. The general accepted birth date of the Monte Carlo methods is (1949) when the first article entitled "The Monte Carlo Methods" by N. Metropolis and S. Ulam appeared in the Journal of the American Statistical Association, 1949 [١٨]. Shortly therefore, Monte Carlo methods used to evaluate complex integrals [٥], and solution of certain differential and integral equations [٤].

The evaluation of definite and multiple integrals is one of the most important fields of applications of Monte Carlo methods. A large no. of deterministic formulas is available for the evaluation of single integrals [٣]. The Monte Carlo methods are not competitive in this case. However, in the case of the multi-dimensional integral, numerical techniques, such as Trapezoidal and Simpson's rules become less attractive. It is more convenient to compute such integrals by Monte Carlo methods which becomes indispensable, which, although less accurate than conventional quadrature formulas, but it is simpler to use [١٦].

The problems handled in this thesis are divided into three chapters, the 1st chapter introduce definitions and some concepts for the simulation and generating random variables. The 2nd discusses the methods of the Monte Carlo simulation for solving the integrals "the Hit or Miss Method and the Sample Mean method" with three sections. The 1st section for the one dimensional integrals and the efficiency between the two methods with examples and. The 2nd extended these methods for the two dimensional integrals and the efficiency between them also with examples. And the last section discusses the solution of the n-dimensional integrals by these methods with the efficiency and examples. Finally the 3rd chapter take four techniques for reducing the variance of the Monte Carlo methods, which are: The importance sampling, Correlated coefficient, Partition of the region, and Biased estimator

List of Contents

List of Tables

Notes and Abbreviations

r.v ≡ *random variable*

r.s ≡ *random sample*

S.S. ≡ *sample space*

eq. ≡ *equation*

func. ≡ *function*

disc. ≡ *discrete*

cont. ≡ *continuous*

distn. ≡ *distribution*

p.d.f ≡ *probability density function*

c.d.f ≡ *cumulative distribution function*

m.g.f ≡ *moment generating function*

no. ≡ *number*

e.w ≡ *else were*

s.t ≡ *such that*

w.r.t ≡ *with respect to*

sec. ≡ *section*

seq. ≡ *sequence*

eff. ≡ *efficiency*

prob.	≡	*probability*
C.I	≡	*confidence interval*
C.L.T	≡	*Central Limit Theorem*
i.i.d	≡	*identically independent distribution*
I.T	≡	*inverse transform*
AR	≡	*acceptance-rejection*
HM	≡	*hit or miss*
SM	≡	*sample mean*
IS	≡	*importance sampling*
PR	≡	*partition of the region*
BE	≡	*biased estimator*
M.S.E	≡	*Mean Square Error*

References

1. Alexander M. Mood, Franklin A. Graybill, Duane C. Boes, Introduction to the theory of statistics "3rd edition", *McGraw Hill, London*, 1974.
2. Averill M. Law & W. David Kelton, Simulation Modeling and Analysis "3rd edition", *McGraw Hill*, 2000.
3. Davis, P. J., Numerical Integration, *Blasdell Publishing Company*, 1967.
4. Fadhel S., On Monte Carlo Methods, M.Sc. thesis, *Saddam University*, 1993.
5. Grant D., Vora, G. and Weeks D., Path-Dependent Options: Extending the Monte Carlo Simulation Approach, *Management Science, Vol. 43, No. 11, pp. 1589-1602*, 1997.
6. Hogg R.V. and Graig A.T., Introduction to Mathematical Statistics "4th Edition", *Macmillian Poblishing Co*., Inc., New York, 1978.
7. Hull, T.E. and A.R. Dobell, Mixed Congruential random number generators for binary machines, *J. Assoc. Comp. Mach.*, 11, 1964, 31-40.
8. Knuth, D. E., The Art of Computer Programming: *Semi numerical Algorithm*, Vol.2, Addision-Wesely, Reading, Massachusetts, 1969.
9. Lehmer, D. H., Mathematical Methods in large-scale computing units, *Ann. Comp. Lab. Harvard Univ.*, 26, 141-146, 1951.
10. MacLaren, M. D. and G. Marsaglia, Uniform random number generators, *J. Assoc. Comp. Mach.*, 12, 1965, 83-89.

11. Marshall, A. W., The use of multi-stage sampling schemes in Monte Carlo computations. in *Symposium on Monte Carlo Methods*, edited by M. A. Meyer, Wiley, New York, 1956, pp. 123-140.

12. Olmstead, P. S., Distribution of sample arrangements for run up and down, *Ann. Math. Stat.*, 17, 1964, 24-33.

13. Powell, M. I. D. and I. Swann, Weighted uniform sampling – A Monte Carlo technique for reducing variance, *J. Inst. Math. App.*, 2, 1966, 228-238.

14. Rand Corporation, *A Million digits with 1000,000 Normal Deviates*, Free press, Clencoe, Illinios, 1955.

15. Reuven Y. Rubinstein, Simulation and the Monte Carlo method, *John Wiley & Sons*, New York, 1981.

16. Shreider, Yu. A. (Ed.), *The Monte Carlo Method (the Method of Statistical Trials)*, Pergamon, Elmsford, New York, 1966.

17. Sobol, I. M., The Monte Carlo method, *University of Chicago*, 1974.

18. Van Neumann, J., Various Techniques used in Connection with Random Digits, *U.S. Bur. Stand. Appl. Math. Ser.,* No. 12, pp. 36-38, 1951.

19. Woller, The Basics of Monte Carlo Simulation, *T. J.*, 1996.

الأهداء

إلى من أستحق امتناني...

إلى من حرّرني فأحياني...

عرفاناً بالجميل...

حرّيتي حياتي

أكرم الصباغ
٢٠٠٨/٤/٩

قال الله تعالى في كتابه الكريم

صدق الله خالقنا العظيم

الخـــلاصـــــة

تطرّقنا في هذه الرسالة إلى طريقتين من طرائق مـــونـــت كـــارلـــو لإيجاد تكاملات المتعددة. الخواص الإحصائية لهذه الطرائق وضِّحت ووحِدت. إن الحد الأدنى للقيمة العليا لمحاولات تخمين هذه التكاملات، فترات الثقة، و الكفاءة لكل طريقة تم اشتقاقها نظرياً و اختبرت عملياً. التكاملات مع الضوضاء تم إيجادها للتكاملات الأحادية فقط. تخفيض التباين لطرائق مونت كارلو نوقِشَت نظرياً و دعّمت بالخوارزميّات باستخدام أربع تقنيّات، سميّت:

“The Importance Sampling, The Correlated Sampling, The Partition of the Region, and The Biased Estimator”

برامج الحاسوب وضِعت في ملحقات ونفِذت باستخدام برنامج

"MathCAD 2001i"