A RANDOM-DISCRETIZATION BASED MONTE CARLO METHOD FOR NUMERICAL INTEGRATION

By

Genghui Wu

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
AT
UNIVERSITY OF MANITOBA
WINNIPEG MANITOBA
AUGUST 2003

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

A RANDOM-DISCRETIZATION BASED MONTE CARLO METHOD FOR NUMERICAL INTEGRATION

BY

Genghui Wu

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree of

Master of Science

UNIVERSITY OF MANITOBA

Date: **August 2003**

Author: **Genghui Wu**

Title: **A Random-Discretization Based Monte Carlo Method for Numerical Integration**

Department: **Statistics**

Degree: **M.Sc.** Convocation: **October** Year: **2003**

Table of Contents

List of Tables

List of Figures

Abstract

Importance sampling is a popular tool for numerical integration. The accuracy of its result, however, depends on the chosen importance density — a good choice may increase efficiency, but an inappropriate choice can cause devastating loss in accuracy. Moreover, in practice there is no guaranteed way to choose a good importance density. This problem becomes more severe when the dimension of the integral increases.

Recently, Fu and Wang (2002) have developed a new Monte Carlo method for multivariate sampling. In this paper, we extend their algorithm to numerical integration problems. We demonstrate that this approach has many advantages over importance sampling. In particular, this algorithm is dimension-free and easy to implement. It is also applicable to many kinds of real functions arising in real life, including non-differentiable, and discontinuous functions. Some benchmark examples are used to demonstrate this algorithm and to compare it with the importance sampling method.

Acknowledgements

We would like to thank Dr. Liqun Wang, my supervisor, for his many suggestions and constant support during this research.

I am also thankful to Dr. James Fu, Dr. Jeffrey Pai and Dr. Dannis Murphy who shared with me their knowledge and ideas, and provided many useful references and friendly encouragements.

The Natural Sciences and Engineering Research Council of Canada (NSERC) scholarship awarded to me was crucial to the successful completion of this thesis.

Of course, I am grateful to my parents for their patience, love, and support.

I am also thankful to my girl friend, Ms. Ling Gu, for her forever love.

Winnipeg, Manitoba — Genghui Wu

August 22, 2003

Introduction

A variety of statistical problems result in evaluating integrals of the following form:

$$\mathrm{I} = \int_{\mathbb{R}^d} g(x)\, dx, \tag{0.0.1}$$

where $d \geqslant 1$, $x = (x_1, x_2, \cdots, x_d) \in \mathbb{R}^d$, and $g \in \mathrm{L}^1(\mathbb{R}^d)$, which is the set of all integrable functions on $\mathbb{R}^d$. For example, in many statistical analysis, researchers want to calculate the moments of certain random variables, such as posterior means, variances of a Bayesian posterior distributions.

The development of a method of calculating integrals of the form (0.0.1) comes to life and becomes useful when one recognizes that many quantities of interest may be cast as expectations with respect to certain probability measures. For example, it is possible to express all probabilities, integrals, and summations as expectations.

Probabilities: Let Y be a random variable, the probability that Y takes on some value in a set A can be expressed as an expectation using the indicator function:

$$P(Y \in A) = \mathrm{E}[I_A(Y)], \tag{0.0.2}$$

where $I_A(Y)$ is the indicator function that takes the value 1 when $Y \in A$, and 0 when $Y \notin A$.

Integrals: Consider a problem now which is completely deterministic—integrating a function $q(x)$ from a to b. So we have $\int_a^b q(x)\,dx$. This can be expressed as an expectation with respect to a uniformly distributed, continuous random variable U between a and b. U has density function $f_U(u) = 1/(b-a)$, so if we rewrite the integral we get

$$\begin{aligned}(b-a)\int_a^b q(x)\frac{1}{b-a}dx &= (b-a)\int_a^b q(x)f_U(x)dx \\ &= (b-a)\mathrm{E}[q(U)].\end{aligned} \tag{0.0.3}$$

Discrete Sums: The discrete version of the above is just the sum of a function $q(x)$ over the finite values of x in a set A, where $A = \{x_i, i = 1, 2, \cdots, n\}$. If we have a random variable W, which takes values in A and all with equal probability $p = P(W = w_i)$. Then the sum may be cast as the expectation as

$$\begin{aligned}\sum_A q(x_i) &= \frac{1}{p}\sum_A q(x_i)p \\ &= \frac{1}{p}\mathrm{E}[q(W)].\end{aligned} \tag{0.0.4}$$

The immediate consequence of this is that all probabilities, integrals, and summations can be approximated by the Monte Carlo method. A crucial thing to note, however, is that there is no restriction that says U or W above must have uniform distributions. This is just for easy illustration of the points above. We will explore this point more while considering importance sampling in the later chapters.

Various techniques have been developed for computing integrals. Generally speaking, those techniques can be classified into two categories. The first category contains deterministic methods, such as the methods of asymptotic expansion and multiple quadrature. They are commonly used for the integrals with a certain degree of smoothness and low dimension. A detailed discussion of deterministic methods can be found in Evans and Swartz (2000). The second category includes stochastic methods, such as the methods of Monte Carlo simulation, importance sampling, and the Markov Chain Monte Carlo (MCMC). Generally speaking, the stochastic methods perform better when the integrands are of relatively high dimensions (say > 4) or involve ill-shaped integrands. General discussions of stochastic methods are given in Tijms, H. C.(2003), Evans and Swartz (2000), and the reference therein.

One of the popular stochastic methods for numerical integration is the importance sampling method. It requires specification of a density function $h(x)$ known as an *importance sampler*. Further, the support of $h(x)$ should cover that of $g(x)$; Then,

the integrands in (0.0.1) can be written as

$$\mathrm{I} = \int_{\mathbb{R}^d} \frac{g(x)}{h(x)} h(x)dx = \mathrm{E}_h\Big[\frac{g(X)}{h(X)}\Big], \quad \text{where } X \sim h(x). \tag{0.0.5}$$

That is to say, the original integral can be written as an expectation of h/h with respect to the density $h(x)$. As we will see in later chapters, a well-chosen importance sampler $h(x)$ can lead to greatly improved estimates of an integral compared with those from among the nominal methods. Therefore, importance sampling method is also sometimes viewed as a variance reduction technique.

However, it should be observed that the variance has been reduced at the cost of doubling the number of function evaluations. There are also some restrictions on choosing a proper importance sampler. Even though we follow such criteria to choose an importance sampler, importance sampling approximations can still fail dramatically. One of the main reason is that, if $h(x)$ decreases towards 0 faster than $g^2(x)h^2(x)$ as x moves away from its modes, we will experience an infinite variance (see Chapter 2). Moreover, as the dimension increases, the task of choosing an importance sampler h becomes more difficult to accomplish. An improper choice of importance sampler will only result in an unsatisfactory approximation. Therefore, finding a good importance sampler function is neither an easy nor safe task.

There are many variance reduction techniques that can be used in conjunction with importance sampling, we will briefly discuss an important special case called Adaptive Importance Sampling (AIS). Discussion of a wider class of techniques can

be found in Evans and Swartz (2000), Owen and Zhou (2000), Man-Suk and Berger (1991), etc.

Recently, Fu and Wang (2002) developed an alternative approach to multivariate random sample generation. Their method combines numerical and sampling-based approaches. In particular, it involves analytical approximation of the density function, random discretization and contourization of an empirical space induced by samples from the uniform distribution on [0,1], and sampling observations from contours of the empirical space according to the discretized density function.

In this thesis, we will apply the random-discretization based monte carlo sampling method to find sufficient and accurate importance region, followed by calculation of the sample mean as an estimate of the integral over a determined importance region. We also use the magnitude of the mean square error as a criterion for the goodness of the estimator. We will demonstrate that this approach has several advantages over importance sampling method or adaptive importance sampling method in certain circumstances.

In Chapter 1, we briefly review the traditional monte carlo numerical integration method. Chapter 2 is devoted to importance sampling. Chapter 3 discuss the use of adaptive importance sampling technique for numerical integration. In Chapter 4, we describe the method of random-discretization based monte carlo Sampling over an importance region, and how this method applies to integration problems. Some

benchmark examples are presented in Chapter 5, and finally, we conclude with some discussion on further research in Chapter 6 .

Chapter 1

Monte Carlo Method for Numerical Integration

In this Chapter we will have a brief look at Monte Carlo Sampling method for numerical integration.

1.1 Monte Carlo Methods

The term "Monte Carlo" was apparently first used by Ulam and von Neumann as a Los Alamos code word for stochastic simulations, which applied to build better atomic bombs. Their methods, involving the laws of chance, were aptly named after the international gaming destination; and soon after the War, a wide range of sticky problems yielded to the new techniques. Despite widespread use of the method, and numerous descriptions of them in articles and monographs, it is difficult to find a succinct definition of "Monte Carlo method" in the literature. Perhaps this is owing to the intuitive nature of the topic, which spawns many definitions by way of specific

examples. Some authors prefer to use the term "stochastic simulation" for almost everything, reserving "Monte Carlo" only for Monte Carlo integration and Monte Carlo tests (Ripley 1987). Others seem less concerned about blurring the distinction between simulation studies and Monte Carlo methods.

Monte Carlo is the art of approximating an expectation by the sample mean of a function of simulated random variables; moreover, it invokes the law of large numbers to approximate expectations. While most Monte Carlo simulations are done by computer today, there were many applications of Monte Carlo methods using coin-flipping, card-drawing, or needle-tossing (rather than computer generated pseudo-random numbers) as early as the turn of the last century — long before the name Monte Carlo arose.

1.2 Monte Carlo Integration

Consider a random variable (or vector) X with probability mass function or probability density function $f_X(x)$, and $g(x)$ is a real-valued function whose domain includes the possible values of X. Then the expected value of $g(x)$ can be written as (Bain and Engelhardt, 1992 p.72):

$$\mathrm{E}\left[(g(X)\right] = \sum_x g(x) f_X(x) \quad \text{if } X \text{ is discrete,} \tag{1.2.1}$$

$$\mathrm{E}\left[g(X)\right] = \int_{x \in \mathbb{R}^d} g(x) f_X(x) dx \quad \text{if } X \text{ is continuous.} \tag{1.2.2}$$

In this work, we are primarily interested in continuous random variables unless otherwise specified. Now, let us denote $\mathcal{X} = (X_1, \cdots, X_n)$, where the $X_i, i = 1, 2, \cdots, n$ are independently, identically distributed. Then, the sample mean of $g(x)$ is given by

$$\hat{I}_n = \bar{g}_n(\mathcal{X}) = \frac{1}{n}\sum_{i=1}^{n} g(X_i),$$

which is a Monte Carlo estimator of $\mathrm{E}[g(X)]$. One other thing to note at this point is that this estimator, $\bar{g}_n(\mathcal{X})$ is unbiased for $\mathrm{E}[g(X)]$:

$$\begin{aligned} \mathrm{E}[\bar{g}_n(\mathcal{X})] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[g(X_i)] \\ &= \mathrm{E}[g(X)]. \end{aligned} \tag{1.2.3}$$

The accuracy of this estimator $\bar{g}_n(\mathcal{X})$ may be described by its variance:

$$\begin{aligned} \mathrm{Var}[\bar{g}_n(\mathcal{X})] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right] \\ &= \frac{\mathrm{Var}[g(X)]}{n} \\ &= \frac{1}{n}\int_{\mathbb{R}^d}\Big[g(x) - \mathrm{E}[g(X)]\Big]^2 f_X(x)dx, \end{aligned} \tag{1.2.4}$$

provided $E[g^2(x)] < \infty$. So, for any given sample size n, the variance of the estimator $\bar{g}_n(\mathcal{X})$ can be calculated by (1.2.4). As the sample size n increases to infinity, the variance of $\bar{g}_n(\mathcal{X})$ decreases to zero. In fact, if $\mathrm{E}[g(X)]$ exists, then the weak law of large numbers (WLLN) tells us that for any arbitrarily small $\epsilon > 0$,

$$\lim_{n\longrightarrow\infty} P\Big(|\hat{I}_n - \mathrm{E}[g(X)]| \geq \epsilon\Big) = 0. \tag{1.2.5}$$

Furthermore, the strong law of large numbers (SLLN) says

$$P\left(\lim_{n\to\infty} \hat{I}_n = \mathrm{E}[g(X)]\right) = 1. \tag{1.2.6}$$

It guarantees almost certain convergence of the approximation; i.e., as long as n is large enough, $\hat{I}_n$ arising from a Monte Carlo experiment shall be close to $\mathrm{E}[g(X)]$ as desired.

1.3 Random Number Generation

Usually, monte carlo methods require the generation of random samples from a wide variety of probability distributions. Our purpose here is to review the primary methods for constructing generators. The books which deal with random variate generation include Bain Engelhardt (1991), Evans and Swartz (2000), Ross (1997), etc.

1. Uniform random variables generation. All random number generators are based on a procedure of generating a sequence $u_1, u_2, \cdots$ of values in $[0, 1]$ that can be considered a realization of an independent and identically distributed $(i.i.d)$ sequence from a uniform distribution. By this, we mean that the sequence $u_1, u_2, \cdots$ has all the properties that an $i.i.d$ uniform sequence from $Unif(0, 1)$ distribution is supposed to have.

The most common class of generators is based on a choice of a number M, which is big enough, a function $f : \{0, 1, \cdots, M-1\}^k \to \{0, 1, \cdots, M-1\}$ and a set of

initial or random seed values $u_1, u_2, \cdots, u_k$. The sequence is then generated via

$$u_{i+1} = f(u_{i-k+1}, \cdots, u_i), \tag{1.3.1}$$

for $i = k, k+1, \cdots$.

2. Inverse methods for random variable generation. A direct method of generating random numbers from a non-uniform probability distribution, when possible, is to use inversion. Suppose that we have a random variable X with distribution function F. Then for $u \in [0,1]$, define the inverse distribution function $F^{-1} : [0,1] \to \mathbb{R}$ by

$$F^{-1}(u) = \inf\{x | u \leqslant F(x)\}.$$

We note that

(i) F^{-1} is increasing,

(ii) $F\Big(F^{-1}(u)\Big) \geqslant u$, and $F^{-1}\Big(F(x)\Big) \leqslant x$,

(iii) $u \leqslant F(x)$ if and only if $F^{-1}(u) \leqslant x$,

(iv) F^{-1} is left-continuous.

We then have the following result

Lemma 1.3.1. *If* $U \sim Unif(0,1)$, *then* $X = F^{-1}(U) \sim F$.

3. Rejection Method for random variable generation. Suppose that f is a (possibly unnormalized) density on $\mathbb{R}$ from which we wish to sample, g is a density on

$\mathbb{R}$ and there exists $c > 0$ such that $f \leqslant cg$. Suppose further we have an algorithm to generate random numbers from g. The rejection algorithm then proceeds as follows:

1. Generate $Y \sim g$ and independently generate $U \sim Unif(0, 1)$;

2. If $Ucg(Y) > f(Y)$, then, reject the generated Y and repeat step 1. Otherwise, return $X = Y$ and stop.

Of course, there are many other methods for generating random variables, such as adaptive rejection method, hit-or-miss method, Gibbs Sampling and more generally, the MCMC methods.

Chapter 2

Importance Sampling

Though it is typically easy to formulate a quantity as an expectation and to propose a "naive" Monte Carlo estimator, it is quite another thing to actually have the Monte Carlo estimator provide you with good estimates in a reasonable amount of computer time. For most problems, a number of Monte Carlo estimators may be proposed; however, some Monte Carlo estimators are clearly better than others. Typically, a "better" Monte Carlo estimator has smaller mean square error (MSE) for the same amount of computational effort than its competitors.

To introduce importance sampling we consider its opposite. Suppose we want a Monte Carlo approximation to $\int_0^1 g(x)dx$. Further suppose that $g(x) = 0$ for $x < 0$ and $x > 1$.

If we have $U \backsim Unif(0,1)$, then we can cast the integral as the expectation with respect to U: $\int_0^1 g(x)dx = \mathrm{E}[g(U)]$. So we may approximate it by a Monte Carlo estimator, $\frac{1}{n}\sum_{i=1}^n g(U_i)$ using a random sample $U_1, U_2, \cdots, U_n$. This would work

reasonably well; however, another possibility is one can use $W \backsim Unif(0,5)$ giving $\int_0^1 g(x)dx = 5\mathrm{E}[g(W)]$, hence the Monte Carlo estimator is $\frac{5}{n}\sum_{i=1}^{n} g(W_i)$. Obviously, such a course of action makes no sense at all because, on average, 80% of the realized w_i's would tell you nothing substantial about the integral of $g(x)$ since $g(x) = 0$ for $1 < x < 5$. is sometimes called "barely relevant sampling". The variance of the estimator is:

$$\begin{aligned}
\mathrm{Var}\Big[\frac{5}{n}\sum_{i=1}^{n} g(W_i)\Big] &= \frac{25}{n}\mathrm{Var}[g(W)] \\
&= \frac{25}{n}\Big[\frac{1}{5}\int_0^5 g^2(w)dw - (\int_0^5 \frac{1}{5}g(w)dw)^2\Big] \\
&= \frac{1}{n}\Big[4\int_0^1 g^2(w)dw + \mathrm{Var}[g(U)]\Big] \\
&= \frac{4}{n}\int_0^1 g^2(w)\,dw + \mathrm{Var}\Big[\frac{1}{n}\sum_{i=1}^{n} g(U_i)\Big] \\
&> \mathrm{Var}\Big[\frac{1}{n}\sum_{i=1}^{n} g(U_i)\Big]
\end{aligned}$$

It does make clear that one's choice of distribution from which to draw random variables will affect the quality of the Monte Carlo estimator.

In this Chapter we will introduce the technique of importance sampling for approximating the integral in (0.0.1).

2.1 Fundamental Theories of Importance Sampling

Definition 2.1.1. If $f(x)$ is a density function of X, then the support of $f(x)$ is denoted as $Supp(f)$, i.e.,

$$Supp(f) = \{x \in \mathbb{R}^d : f(x) \neq 0\}. \tag{2.1.1}$$

Consider the approximation of the integral

$$I(g) = \int_{\mathbb{R}^d} g(x)dx, \tag{2.1.2}$$

where $x \in \mathbb{R}^d$, and $g \in L^1(\mathbb{R}^d)$ is integrable. The method of importance sampling requires specification of a density function $h(x)$, known as an *importance sampler*, that satisfies $Supp(g) \subseteq Supp(h)$. We can then write (2.1.2) as

$$I(g) = \int_{\mathbb{R}^d} \frac{g(x)}{h(x)} h(x)dx = \mathrm{E}_h\Big[\frac{g(X)}{h(X)}\Big], \tag{2.1.3}$$

where $X \sim h(x)$. The original integral is written as an expectation with respect to the density $h(x)$. As an additional requirement, we suppose that we have a reasonably efficient algorithm for generating random samples from $h(x)$. Then based on a sample $X_1, \cdots, X_n$ from $h(x)$, the importance sampling estimator of $I(g)$ is given by

$$\hat{I}_{h,n} = \frac{1}{n}\sum_{i=1}^{n} \frac{g(X_i)}{h(X_i)}. \tag{2.1.4}$$

A justification for (2.1.4) to be a valid estimator for $I(g)$ in (2.1.3) is the strong law of large numbers, which implies that

$$P[\lim_{n\to\infty} \hat{I}_{h,n} = I(g)] = 1. \tag{2.1.5}$$

Therefore, for large values of n we expect $\hat{I}_{h,n}$ to be a good approximation to $I(g)$. If we are to examine the goodness of the estimator for any given finite sample size n, we use the measure of the mean square error (MSE), defined as following (Bain and Engelhardt 1992, pp.309).

Definition 2.1.2. If T is an estimator of $\tau(\theta)$, where $\tau(\theta)$ is a known function of parameter θ., then the *bias* of T is defined as

$$b(T) = \mathrm{E}_\theta(T) - \tau(\theta), \tag{2.1.6}$$

and the *mean square error* (MSE) of T is defined as

$$\mathrm{MSE}(T) = \mathrm{E}_\theta[T - \tau(\theta)]^2. \tag{2.1.7}$$

Proposition 2.1.1. *If T is an estimator of $\tau(\theta)$, then*

$$\begin{aligned} \mathrm{MSE}(T) &= \mathrm{E}_\theta[T - \tau(\theta)]^2 \\ &= \mathrm{E}_\theta[T - \mathrm{E}_\theta(T)]^2 + [\mathrm{E}_\theta(T) - \tau(\theta)]^2 \\ &= \mathrm{Var}(T) + [b(T)]^2. \end{aligned} \tag{2.1.8}$$

The MSE is a reasonable criterion that considers both the variance and the bias of an estimator. It provides a useful means for comparing two or more estimators. In our case, we have

$$\mathrm{MSE}(\hat{I}_{h,n}) = \mathrm{E}_h\Big[\hat{I}_{h,n} - I(g)\Big]^2.$$

Since $\hat{I}_{h,n}$ is an unbiased estimator of $I(g)$, i.e. $\mathrm{E}(\hat{I}_{h,n}) = I(g)$, then,

$$\begin{aligned} \mathrm{MSE}(\hat{I}_{h,n}) &= \mathrm{Var}_h[\hat{I}_{h,n}] \\ &= \frac{\sigma_h^2}{n}, \end{aligned} \tag{2.1.9}$$

where

$$\begin{aligned} \sigma_h^2 &= \mathrm{Var}_h\Big[\frac{g(X)}{h(X)}\Big] \\ &= \int_{\mathbb{R}^d} \Big[\frac{g(x)}{h(x)} - I(g)\Big]^2 h(x)dx \\ &= \int_{\mathbb{R}^d} \frac{g^2(x)}{h(x)}dx - I^2(g) \\ &< \infty \end{aligned} \tag{2.1.10}$$

provided $g^2(x)/h(x) \in L^1(\mathbb{R})$. Then by the Central Limit Theorem (CLT) (Bain and Engelhardt 1992, p.238), we have the following result.

Proposition 2.1.2. *Suppose that $X_1, X_2, \cdots, X_n$ are i.i.d. random variables from the importance sampler $h(x)$, which satisfies the conditions $\sigma_h^2 < \infty$ and $Supp(g) \subseteq Supp(h)$. Then as $n \longrightarrow \infty$,*

$$\frac{\hat{I}_{h,n} - I(g)}{\sigma_h/\sqrt{n}} \xrightarrow{d} N(0,1). \tag{2.1.11}$$

Definition 2.1.3. For sequences X_n, Y_n of random variables, we write $X_n = O_p(Y_n)$, if for any $\epsilon > 0$, there exist $M_\epsilon > 0$, and $N_\epsilon > 0$ such that

$$P\left(\Big|\frac{X_n}{Y_n}\Big| > M_\epsilon\right) < \epsilon$$

for all $n \geqslant N_\epsilon$. (i.e., the sequence X_n/Y_n is bounded in probability.)

Then proposition 2.1.2 implies that

$$\hat{I}_{h,n} - I(g) = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{2.1.12}$$

where $1/\sqrt{n} \to 0$ as $n \to \infty$. In this case, we say that importance sampling estimator converges to $I(g)$ at the rate $1/\sqrt{n}$. Note that this requires the variance of the estimator to be finite. Also, we observe that the rate of convergence in (2.1.12) is independent of the dimension d.

In practice, the quantity of σ_h^2 in (2.1.9) is unknown. Since

$$s_h^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{g(X_i)}{h(X_i)}\right]^2 - (\hat{I}_{h,n})^2. \tag{2.1.13}$$

By the SLLN, and (2.1.5), we have the following result.

Proposition 2.1.3. *Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. random variables from $h(x)$, which satisfies $\sigma_h^2 < \infty$ and $Supp(h) \subseteq Supp(g)$. Then,*

$$P[\lim_{n\to\infty} s_h^2 = \sigma_h^2] = 1.$$

Furthermore, we have the following result

Proposition 2.1.4. *Suppose that $X_1, X_2, \cdots, X_n$ are i.i.d. random variables from an importance sampler $h(x)$, which satisfies the conditions $\sigma_h^2 < \infty$ and $Supp(g) \subseteq Supp(h)$. Then as $n \longrightarrow \infty$,*

$$\frac{\hat{I}_{h,n} - I(g)}{s_h/\sqrt{n}} \xrightarrow{d} N(0,1). \tag{2.1.14}$$

Proof. By the strong law of large numbers, we have that $s_h^2 \to \sigma_h^2$ with probability 1, as $n \to \infty$. Then (2.1.13) follows from Proposition 2.1.2 and Slutsky's theorem (Bain and Engelhardt, 1992, p.248). □

Result (2.1.4) has the interpretation that the interval $\hat{I}_{h,n} \pm 3s_h/\sqrt{n}$ contains the value $I(g)$ with virtual certainty when n is large. We illustrate the following simple example as an application of importance sampling method.

Example 1. Consider the function $g(x) = 4\sqrt{1-x^2}$, $x \in [0,1]$. Suppose we wish to evaluate $I = \int_0^1 g(x)dx$, which is π.

Let us first use the uniform importance sampler $h(x) = 1$, $x \in [0,1]$. In this case,

$$\hat{I}_{h,n} = \frac{1}{n}\sum_{i=1}^{n} 4\sqrt{1-X_i^2},$$

where the $X_i's$ are i.i.d. $\sim Unif[0,1]$. This estimator is unbiased with variance

$$\begin{aligned} \mathrm{Var}(\hat{I}_{h,n}) &= \frac{1}{n}\left(\int_0^1 g(x)^2dx - I^2\right) \\ &= \frac{1}{n}\left(\int_0^1 16(1-x^2)dx - \pi^2\right) \\ &\approx \frac{0.7968}{n}. \end{aligned} \tag{2.1.15}$$

Now, consider another importance sampler $h(x) = 2-2x$, where $x \in [0,1]$. We can then generate $X_1, \cdots, X_n \sim h(x)$, which yields the importance sampling estimator

$$\hat{I}_{h,n} = \frac{1}{n}\sum_{i=1}^{n} \frac{2\sqrt{1-X_i^2}}{1-X_i},$$

whose variance is

$$\begin{aligned} \mathrm{Var}(\hat{I}_{h,n}) &= \frac{1}{n}\left(\int_0^1 \frac{g(x)^2}{h(x)}\,dx - I^2\right) \\ &= \frac{1}{n}\left(\int_0^1 \frac{16(1-x^2)}{2-2x}dx - \pi^2\right) \\ &= \frac{2.0982}{n}. \end{aligned} \tag{2.1.16}$$

2.2 Choice of Importance Sampler

From Example 1, we clearly see that different choices of importance sampler result in different precision of estimators with respect to the magnitude of the sample variance. Choice of the importance sampler $h(x)$ for $I(g)$ is clearly the most significant consideration for a user of importance sampling. We first carry out some generally accepted qualitative criteria for choosing an importance sampler:

1. $h(x) > 0$ whenever $g(x) \neq 0$;
2. $h(x)$ should be close to being proportional to $|g(x)|$;
3. it should be easy to simulate random samples from $h(x)$;
4. it should be easy to compute the density $h(x)$ for any realizable value x.

From Section 2.1, we want to choose $h(x)$ so that σ_h^2 is as small as possible. The best we can do is to have $\sigma_h^2 = 0$, and this occurs if and only if $P_h(g(X)/h(X) = c) = 1$ for the constant $c = I(g)$. Formally, we state this result as follows.

Proposition 2.2.1. *If* $0 < \int_{\mathbb{R}^d} |g(x)|dx < \infty$, *then the choice of* $h(x)$ *that minimizes*

the variance σ_h^2 *is*

$$h^o(x) = \frac{|g(x)|}{\int_{\mathbb{R}^d} |g(x)|dx},$$

and the minimum variance is

$$\left(\int_{\mathbb{R}^d} |g(x)|dx\right)^2 - \left(\int_{\mathbb{R}^d} g(x)dx\right)^2.$$

Proof. From (2.1.10), we have

$$\begin{aligned}
\sigma_h^2 &= \int_{\mathbb{R}^d} \frac{g(x)^2}{h(x)}\,dx - I(g)^2 \\
&= \left(\int_{\mathbb{R}^d} \mid g(x) \mid dx\right)^2 \int_{\mathbb{R}^d} \frac{h^0(x)^2}{h(x)}dx - I(g)^2 \qquad (2.2.1) \\
&= \left(\int_{\mathbb{R}^d} \mid g(x) \mid dx\right)^2 \Big[\int_{\mathbb{R}^d} \frac{[h^0(x) - h(x)]^2}{h(x)}dx + 1\Big] - I(g)^2
\end{aligned}$$

Hence minimizing σ_h^2 is equivalent to minimizing the chi-square distance

$$\mathrm{E}_h\left[\Big(\frac{h^o(X) - h(X)}{h(X)}\Big)^2\right] = \int_{\mathbb{R}^d} \frac{(h^o(x) - h(x))^2}{h(x)}dx,$$

and this is clearly minimized by taking $h = h^o$. □

From practical point of view, Proposition (2.2.1) is not really of much help in selecting $h(x)$ since it requires knowing the exact value of $\int_{\mathbb{R}^d} |g(x)|dx$ in advance. One usually starts with a certain family of probability densities, such as normal densities, and then tries to find the "optimal" density within this family.

Example 2. Consider once again the function $g(x) = 4\sqrt{1-x^2}$, $x \in [0,1]$ of Example 1. Suppose we wish to evaluate $I = \int_0^1 g(x)dx$. We will choose importance sampler

from the family of linear probability density functions $h(x) = a + bx$, $x \in [0, 1]$. Two members of this family, including the uniform density, have been used in Example 1. We wish to find the best $h^0(x)$, which minimizes the variance (2.1.10).

First, since $\int_0^1 (a + bx)\, dx = 1$, we have

$$a + b \int_0^1 x\, dx = a + \frac{b}{2} = 1$$

which implies $a = 1 - \frac{b}{2}$. Secondly, since $\forall x \in [0, 1]$, $a + bx \geq 0$, it follows $-2 \leq b \leq 2$. Therefore, the family of all linear density functions over $[0, 1]$ is $\{1 - \frac{b}{2} + bx\,, 0 \leqslant x \leqslant 1 : -2 \leqslant b \leqslant 2\}$. Some member of this family are shown graphically in Figure 2.1, together with the integrand $g(x)$. Within this linear importance sampler family, the importance sampling estimator and its variance for Example 1 can be expressed as

$$\hat{I}_{h,n} = \frac{1}{n} \sum_{i=1}^{n} \frac{4\sqrt{1 - X_i^2}}{1 - b/2 + bX_i},$$

and

$$\mathrm{Var}(\hat{I}_{h,n}) = \frac{1}{n} \int_0^1 \frac{16(1 - x^2)}{1 - b/2 + bx} dx - \pi^2.$$

After some reduction, we have

$$\mathrm{Var}(\hat{I}_{h,n}) = \frac{1}{n} \left(\frac{16b - 32}{b^3} + \frac{4}{b^3}[4b^2 - (2 - b)^2] \ln \frac{2 + b}{2 - b} - \pi^2 \right).$$

Table 2.1 contains variances $\mathrm{Var}[\hat{I}_{h,n}]$ for some selected linear densities h. We graph the value of $\mathrm{Var}[\hat{I}_{h,n}]$ as a function of b in Figure 2.2. We also summarize the values of $\mathrm{Var}(\hat{I}_{h,n})$ with respect to b in Table 2.1. We find that the variance is minimized at

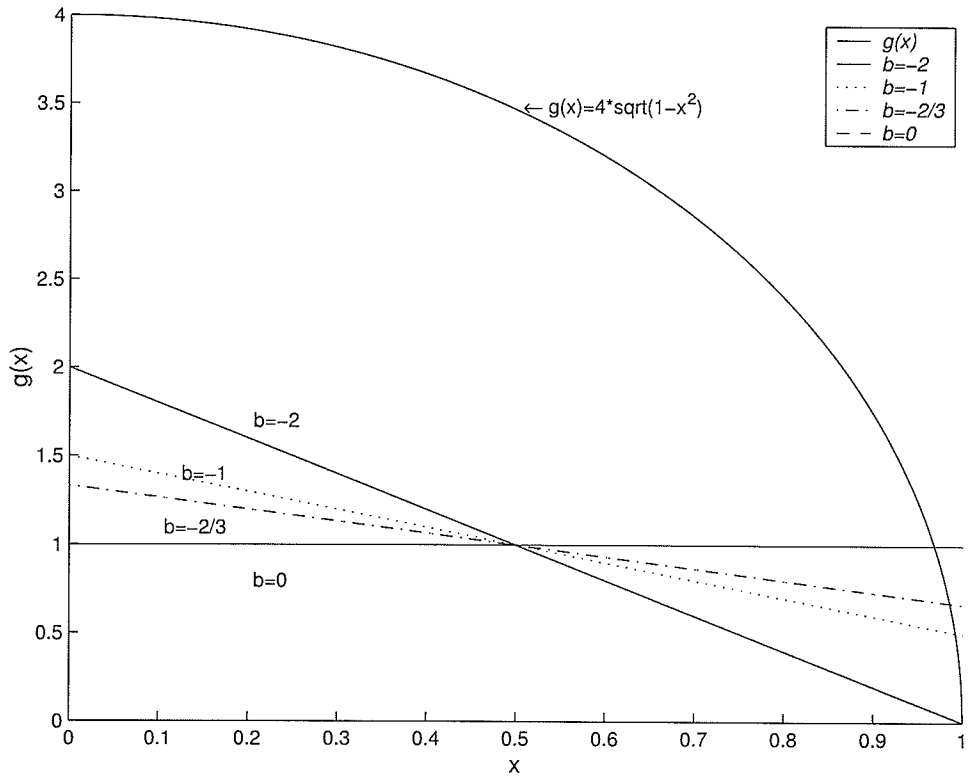

Figure 2.1: Linear importance sampler for $g(x)$ with respect to the value of b

Table 2.1: Variance of Importance Sampling

b	-2	-1	$-2/3$	0
Variance	$2.0982/n$	$0.1581/n$	$0.2240/n$	$0.7968/n$

$b = -1$ with variance of $0.1581/n$, which is a little surprising because the density with $b = -2$ seems to be a better importance sampler than that with $b = -1$ from Figure 2.1. Example 2 shows that it is hard to tell ahead of time whether a given $h(x)$ will reduce the variance or not. Moreover, it is not trivial to find the "optimal" importance sampler even if the density family is given. We see that importance sampling can fail dramatically, even when $h(x)$ is proportional to $g(x)$. When the importance sampler $h(x)$ decreases toward 0 faster than $g^2(x)$ as x moves away from its mode(s), we find that $\mathrm{Var}(\hat{I}_{h,n})$ will inflate rapidly. The irony is that this large variance may be due to a region of $\mathbb{R}^d$ that is unimportant in ordinary Monte Carlo sampling.

For multi-dimensional $g(x)$, it is even harder to find a good importance sampler. As a matter of fact, there are even not many choices of families of distributions as multivariate importance samplers. A commonly used one is the family of multivariate student t distribution (Evans and Swartz 1995, 2000). However, it can be difficult to generate random variables from these densities and have to fit them to a multi-dimensional $g(x)$.

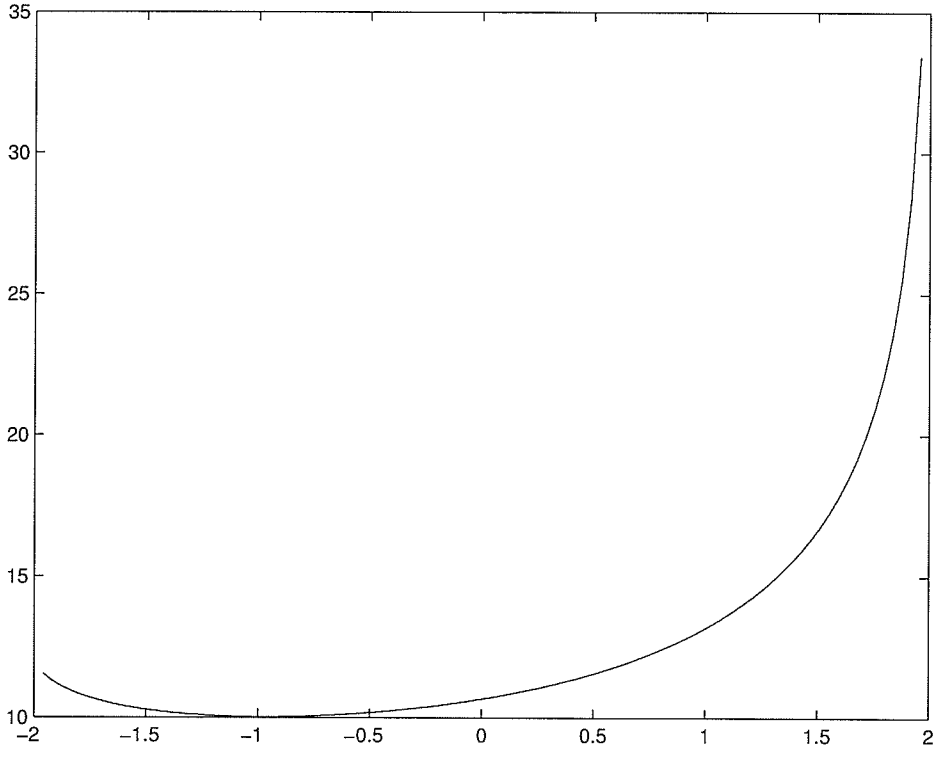

Figure 2.2: Variance of $\hat{I}_{h,n}$ as a function of b, Example 2

Chapter 3

Adaptive Importance Sampling

As stated in the Chapter 2, a main problem in importance sampling is to find a "good" importance sampler. A good importance sampler is one that allows efficient generation of variates, and gives acceptable estimates within practical computing times. Adaptive importance sampling is developed as a mechanical way of dealing with the selection of a good importance sampler within a chosen density family $H = \{h_\lambda : \lambda \in \Lambda\}$. It starts with choosing an arbitrary $h_\lambda \in H$, and runs importance sampling in an iterative way to continuously update λ. The process stops when the desired accuracy is reached.

3.1 Criterion of Choosing Appropriate Importance Sampler Though AIS

Suppose one wish to evaluate the integral:

$$I(g) = \int_{\mathbb{R}^d} g(x)\, dx,$$

where $x \in \mathbb{R}^d$, and $g \in L^1(\mathbb{R}^d)$. The method of importance sampling requires specification of a importance sampler $h_\lambda \in H$. Then, from (2.1.10), one has

$$\begin{aligned} \mathrm{Var}[\hat{I}_{h_\lambda,n}] &= \frac{\sigma^2_{h_\lambda,n}}{n} \\ &= \frac{1}{n}\mathrm{Var}\Big[\frac{g(X)}{h_\lambda(X)}\Big] \\ &= \frac{1}{n}\Big[\int_{\mathbb{R}^d} \frac{g^2(x)}{h_\lambda(x)}\, dx - I^2(g)\Big], \end{aligned} \tag{3.1.1}$$

where $X \sim h_\lambda$.

One's objective is to find an accurate estimator of $I(g)$, $\hat{I}_{h_\lambda,n}$ by minimizing $\mathrm{Var}[\hat{I}_{h_\lambda,n}]$. Since $I^2(g)$ is a fixed value, then minimizing $\mathrm{Var}[\hat{I}_{h_\lambda,n}]$ is equivalent to minimizing $\int_{\mathbb{R}^d} \frac{g^2(x)}{h_\lambda(x)} dx$. From Section 2.2, we know that a good choice of importance sampler $h_\lambda(x)$ should be a mimic of $g(x)$. Here, we will achieve this goal by matching the moments of $h_\lambda(x)$ to the moments of $g(x)$ as close as possible through adaptive importance sampling (Evans and Swartz, 2000).

Suppose g is nonnegative and integrable, which means $g \geqslant 0$, and $0 < \int_{\mathbb{R}^d} g(x)\, dx < \infty$. Then, $\tilde{g}(x) = \frac{g(x)}{\int_{\mathbb{R}^d} g(x) dx}$ is a density function. To illustrate the AIS, let

$V = (X, X^2)'$, then, $E_{\tilde{g}}(V)$ is the vector of the first and second moments of $\tilde{g}(x)$ as following:

$$\begin{aligned} \mathrm{E}_{\tilde{g}}(V) &= \begin{bmatrix} \int_{\mathbb{R}^d} x\, \tilde{g}(x)dx \\ \int_{\mathbb{R}^d} x^2\, \tilde{g}(x)dx \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{E}_{\tilde{g}}[X] \\ \mathrm{E}_{\tilde{g}}[X^2] \end{bmatrix}, \end{aligned} \tag{3.1.2}$$

where $X \sim \tilde{g}(x)$. Next, one wishes to find $h_\lambda \in H$ that minimizes the Euclidean norm as:

$$\| \mathrm{E}_{\tilde{g}}(V) - \mathrm{E}_{h_\lambda}(V) \| = \Big[\mathrm{E}_{\tilde{g}}(V) - \mathrm{E}_{h_\lambda}(V)\Big]' \Big[\mathrm{E}_{\tilde{g}}(V) - \mathrm{E}_{h_\lambda}(V)\Big], \tag{3.1.3}$$

where

$$\begin{aligned} \mathrm{E}_{h_\lambda}(V) &= \begin{bmatrix} \int_{\mathbb{R}^d} x\, h_\lambda(x)dx \\ \int_{\mathbb{R}^d} x^2\, h_\lambda(x)dx \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{E}_{h_\lambda}[X] \\ \mathrm{E}_{h_\lambda}[X^2] \end{bmatrix} \qquad (3.1.4) \\ &= \begin{bmatrix} \mu'_{h_\lambda} \\ \mu''_{h_\lambda}, \end{bmatrix} \qquad (3.1.5) \end{aligned}$$

where $X \sim h_\lambda(x)$.

Therefore, one has the criteria (3.1.3) to measure the goodness of a particular importance sampler h_λ. The smaller value of the Euclidean norm, the better choice of the importance sampler. The iterative process of minimizing Euclidean norm ensures us to find an appropriate importance sampler $h_\lambda \in H$, and guarantees the first two moments of $h_\lambda(x)$ are closest to those of $\tilde{g}(x)$ within the density family H.

3.2 Algorithm of Adaptive Importance Sampling

Suppose $H = \{h_\lambda : \lambda \in \Lambda\}$ is chosen to be the family of potential importance samplers for the given numerical integration problem in (2.1.2). In this Section, an approach of finding an appropriate importance sampler $h_\lambda \in H$ through AIS is introduced.

Step 1: One arbitrarily picks $\lambda_1 \in \Lambda$ and generate $X_1, \cdots, X_n$ from h_{λ_1}.

Step 2: One estimates $\mathrm{E}_{\tilde{g}}(V)$ by

$$\hat{\mathrm{E}}_{\lambda_1}(V) = \hat{I}_{h_{\lambda_1},n}\Big(\frac{V\tilde{g}}{h_{\lambda_1}}\Big),$$

where by (2.1.4)

$$\begin{aligned} \hat{I}_{h_{\lambda_1},n}\Big(\frac{V\tilde{g}}{h_{\lambda_1}}\Big) &= \frac{1}{n}\sum_{i=1}^{n}\frac{V_i\tilde{g}(X_i)}{h_{\lambda_1}(X_i)} \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{g}(X_i)}{h_{\lambda_1}(X_i)}\begin{bmatrix} X_i \\ X_i^2 \end{bmatrix} \\ &= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \dfrac{X_i\,\tilde{g}(X_i)}{h_{\lambda_1}(X_i)} \\ \dfrac{X_i^2\,\tilde{g}(X_i)}{h_{\lambda_1}(X_i)} \end{bmatrix}; \end{aligned} \tag{3.2.1}$$

Step 3: One finds $\lambda_2 \in \Lambda$ that minimizes

$$\| \hat{E}_{\lambda_1}(V) - \mathrm{E}_{h_\lambda}(V) \| = \sqrt{\Big[\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{X_i\,\tilde{g}(X_i)}{h_{\lambda_1}(X_i)} - \mu'_{h_\lambda}\Big)\Big]^2 + \Big[\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{X_i^2\,\tilde{g}(X_i)}{h_{\lambda_1}(X_i)} - \mu''_{h_\lambda}\Big)\Big]^2}.$$

Step 4: Once λ_2 is found, one generates sample $X_{n+1}, \cdots, X_{2n}$ from h_{λ_2}, and estimate $\mathrm{E}_{\tilde{g}}(V)$ by

$$\hat{\mathrm{E}}_{\lambda_2}(V) = \frac{\hat{I}_{h_{\lambda_1},n}\left(\frac{V\tilde{g}}{h_{\lambda_1}}\right) + \hat{I}_{h_{\lambda_2},n}\left(\frac{V\tilde{g}}{h_{\lambda_2}}\right)}{2},$$

Step 5: Similar to step 4, one finds $\lambda_3 \in \Lambda$ that minimizes

$$\| \hat{\mathrm{E}}_{\lambda_2}(V) - \mathrm{E}_{h_\lambda}(V) \| .$$

Finally, this iterative process stops when the estimates $|\hat{\mathrm{E}}_{\lambda_i}(V) - \hat{\mathrm{E}}_{\lambda_{i-1}}(V)| < \epsilon$, where $\epsilon > 0$ is a predetermined number.

In general, $\mathrm{E}_{\tilde{g}}(V)$ can contain higher moments and even probability of a particular subset of $\mathbb{R}^d$ under $\tilde{g}$. Furthermore, we have the following general result for the adaptive importance sampling (Evans and Swartz, 1998, p.211).

Theorem 3.2.1. *If for the adaptive importance sampling procedure,*

$$E\left(\sum_{j=1}^{\infty} \frac{1}{j^2} E_{h_{\lambda_j}}\left[\left(\frac{g(X) - h_{\lambda_j}(X)}{h_{\lambda_j}(X)}\right)^2\right]\right) < \infty,$$

then, $\hat{E}_{\lambda_i}(V)) \to E_{\tilde{g}}(V)$ *almost surely as* $i \to \infty$.

The analysis of convergence issues for adaptive importance sampling is more difficult than importance sampling because of the lack of independence between iterations. However, we have the following conditions and theorem to guarantee the almost sure consistency of this estimate (Evans and Swartz, 1998, p.213).

Theorem 3.2.2. *For the adaptive importance sampling procedure, assume the following conditions*

(i) for each $r \in R$ *there is a unique* $\alpha(r) \in \Lambda$ *such that*

$$\| E_{h_{\alpha(r)}}(V) - r \| = \inf \| E_{h_\lambda} - r \|;$$

(ii) $\alpha : R \to \Lambda$ *is continuous;*

(iii) $\int g(x)/h_\lambda(x)dx$ *is continuous in* λ *for* g^2, Vg^2, *and* V^2g^2;

(iv) $\int g(x)/h_\lambda^2(x)dx$ *is continuous in* λ *for* g^3, $|V|g^3$, V^2g^3, *and* $|V|^3g^3$;

(v) $E\{\sum_{j=1}^{\infty} \frac{1}{j^2} E_{h_{\lambda_j}}[(\frac{g - h_{\lambda_j}}{h_{\lambda_j}})^4]\} < \infty$,

then $\lambda_i = \alpha(\hat{E}_{\lambda_i}(V))) \to \alpha(Eg(V)) = \lambda$ *almost surely as* $i \to \infty$. *Furthermore,* $Var[\hat{I}_{h_\lambda,n}] \to Var[\hat{I}_n]$ *almost surely as* $i \to \infty$.

In the rest of this section, we carry out an example to show that the method of adaptive importance sampling does improve the accuracy of integration if a importance density family set is properly chosen.

Example 3. We consider a simple two dimensional problem. This problem was studied by OH and Berger (1992). Suppose we have

$$g(x) = 0.25g_1(x) + 0.75g_2(x), \tag{3.2.2}$$

where $x = (x_1, x_2)' \in \mathbb{R}^2$, and g_1, g_2 are the density functions of $\mathrm{N}(\mu_1, \Sigma_1)$ and $\mathrm{N}(\mu_2, \Sigma_2)$ respectively, where

$$\mu_1 = \begin{pmatrix} 0.0 \\ \\ 0.0 \end{pmatrix}, \qquad \mu_2 = \begin{pmatrix} 2.1 \\ \\ 2.1 \end{pmatrix},$$

and

$$\Sigma_1 = \begin{pmatrix} 1.0 & 0.8 \\ \\ 0.8 & 1.0 \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} 1.0 & 0.0 \\ \\ 0.0 & 1.0 \end{pmatrix}.$$

Figure 3.1 shows the 2-dimensional density and contour graph of $g(x)$ over $[-6, 6]^2$. It is easy to see that $g(x) \approx 0$ outside the region $[-6, 6]^2$. Suppose that one desires to find the mean μ of g(x) using AIS method. Further suppose $V = (X_1, X_2)'$. Then,

$$\mathrm{E}(V) = \mu = \begin{bmatrix} \mathrm{E}(X_1) \\ \\ \mathrm{E}(X_2) \end{bmatrix}, \tag{3.2.3}$$

where X_1, X_2 are the 1st and 2nd elements of X, respectively. As shown in Figure 3.1, g is skewed in the direction of $(0, 0)'$. Since the skewness is not severe, a multivariate t family, denoted as T, can be chosen to be the importance sampler family because of its thick tails and simplicity in random variate generation.

Following the algorithm described in Section 3.1, AIS continued in this example to the 320th stage, at which point $\hat{\mu}$ was obtained from $\hat{\mathcal{A}}_{320}(V)$. Table 3.1 compares the numerical results produced from using adaptive importance sampling method (AIS)

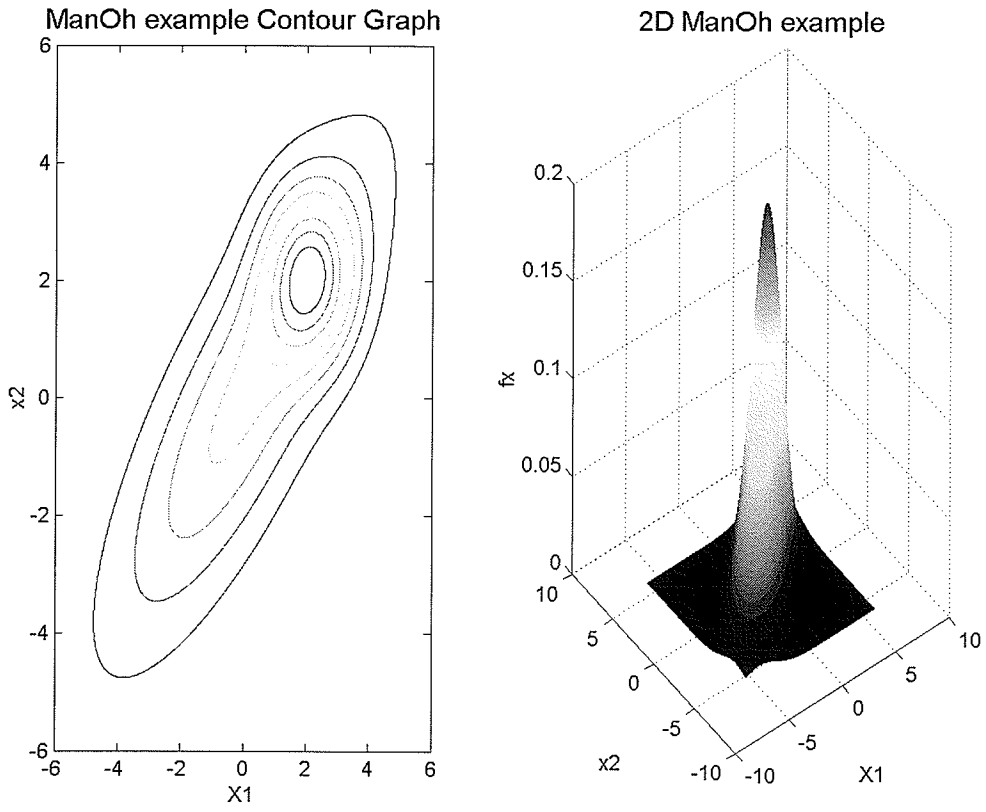

Figure 3.1: 2-D Normal distribution contour and density graphs

to those from using importance sampling method (IS) with one particular importance sampler chosen from T. $\hat{\mu}_1$ and $\hat{\mu}_2$ are estimates of the 1st and 2nd elements of $\hat{\mu}$, respectively.

Table 3.1: Comparison of IS and AIS

Parameter	IS	AIS	True Values
n	57,600	33,000	
$\hat{\mu}_1$	1.5934	1.5837	1.5750
$\hat{\mu}_2$	1.5986	1.5953	1.5750
Relative Error	0.0133	0.0092	
CPU Time(s)	295.0	169.3	

3.3 Limitation of the Adaptive Importance Sampling

Although normally AIS produces a better estimate than does the importance sampling (Example 3), it requires more calculation due to its iterative algorithm specified in Section 3.1, which may entail greater consumption of computer memory and computing time.

In addition, this method requires that we pre-determine a family of importance densities. Hence, it limits our choice of importance samplers. To show this point more intuitively, we consider the following example.

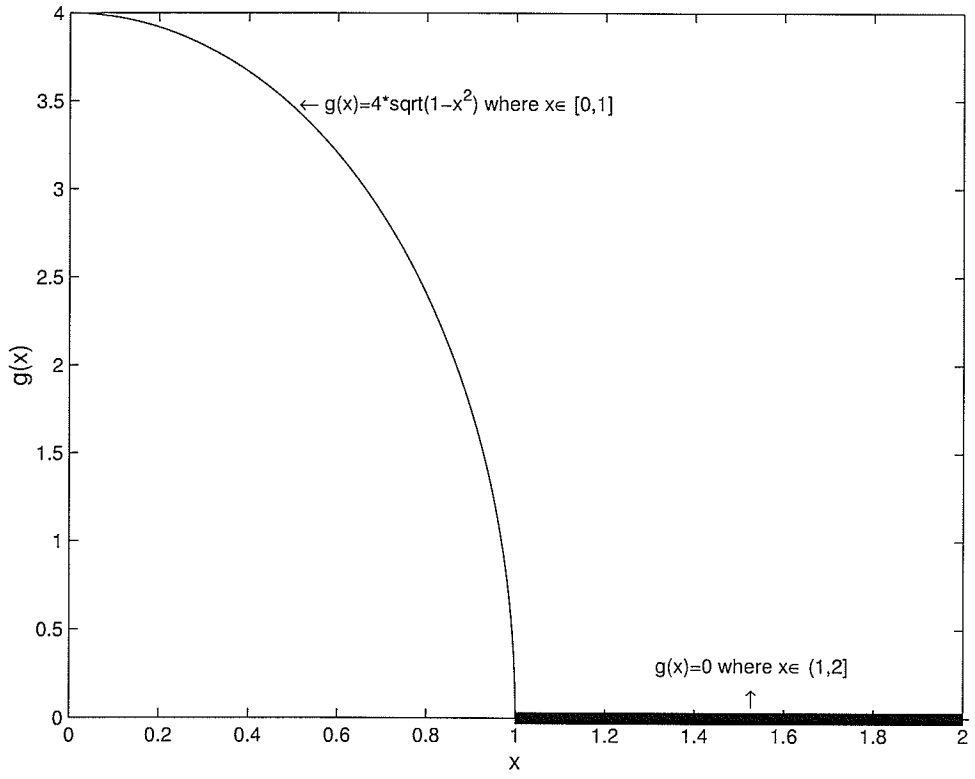

Figure 3.2: Distribution of $g(x)$ in example 4

Example 4.

$$g(x) = \begin{cases} 4\sqrt{1-x^2} & x \in [0,1], \\ 0 & x \in (1,2]. \end{cases}$$

The value of this integral over $[0,2]$ is the same as that considered in Example 1, which is π. The graph of $g(x)$ is shown in Figure 3.2. We consider the linear p.d.f. family $h(x) = a + bx$, $x \in [0,2]$. We want to find the best $h(x)$ which minimizes the sampling variance. Since $\int_0^2 (a+bx)dx = 1$, we have

$$a + b\int_0^2 x\,dx = 2a + 2b = 1,$$

which implies $a = 1/2 - b$. We also know that $a + bx \geq 0$ for all $x \in [0, 2]$, which implies $-0.5 \leq b \leq 0.5$. Then, the variance of the importance sampling estimator based on h is

$$\mathrm{Var}\left[\hat{I}_{h,n}\right] = \frac{1}{n}\int_0^1 \left(\frac{16(1-x^2)}{\frac{1}{2} - b + bx}dx - \pi^2\right).$$

After manipulation, we have

$$\mathrm{Var}(\hat{I}_{h,n}) = \frac{1}{n}\left(\frac{8-24b}{b^2} + \frac{16b-4}{b^3}\ln\frac{1}{1-2b} - \pi^2\right).$$

Figure 3.3 shows this variance as a function of b. We find that the variance is minimized when b at $-1/2$ with a variance of $3.5883/n$. (see Figure 3.3)

We can clearly see that even if we are able to find the best importance sampler, $h_\lambda = 1 - 1/2\,x$ via adaptive importance sampling method in this Example. The negative impact of the unimportant region, $x \in [1, 2]$ will cause variance of the estimator to be much larger than that of the optimal estimator found in Example 1 with respect to the importance region $[0, 1]$, where $b = -1$. In fact, the variance is even larger than the estimator obtained by setting $b = 0$, which corresponds to uniform sampling from the important region $[0, 1]$.

Beyond the difficulty of importance sampling, restricting the search of the importance sampler within a prescribed density family provides no guarantee that it will result in precise approximation of $I(g)$. Especially when we don't know the shape of

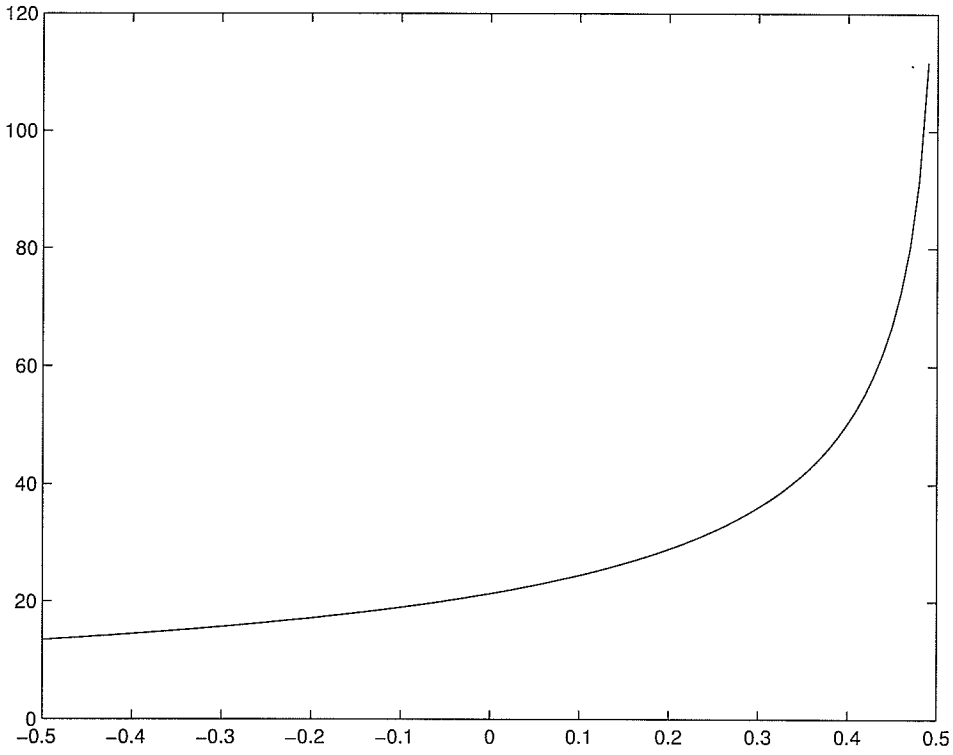

Figure 3.3: Variance of $\hat{I}_{h,n}$ as a function of b, Example 4.

the integrand in a multi-dimensional space, it is quite possible that we will choose a bad density family. This will generally lead to an inaccurate estimation.

Chapter 4

Sampling Over Importance Regions

Fu and Wang (2002) have developed a new Monte Carlo method for multi-variate sampling [9]. In this Chapter. we describe how to apply this method to determine an importance sampling region. We demonstrate the advantages of this approach over both importance sampling and adaptive importance sampling by sampling from a computed optimal importance region.

From above Examples 1 and 4, we can see that it is difficult to find an optimal importance sampler to accurately compute an integral. It requires knowledge of the shape of the integrand $g(x)$. Even if we know the shape of the integrand, as in Example 1, we are still unable to find the best importance sampler without going through complicated numerical calculations.

When $g(x)$ is multi-dimensional, finding a good numerical scheme is obviously

becomes a difficult task. In this Chapter, we approach the multi-dimensional integration problem via importance sampling with a uniform importance sampler within a computed importance region, and we will show this method has many advantages over traditional importance sampling:

1. It is applicable to the integration problems of moderate to high dimensions.
2. It is computationally efficient.
3. It provides a reasonably accurate approximation to $I(g)$.

4.1 Algorithm for Sampling Over a Bounded Support

Consider an integral

$$I = \int_S g(x)\, dx,$$

where $S \subseteq \mathbb{R}^d$ is a bounded subset. Suppose $g(x)$ is non-negative on S, so that it can be regarded as a density function up to a normalizing constant. We also define the *negligible region* on S for evaluating I as the region where $g(x)$ is approximately 0. Our objective is to determine an appropriate sampling region, which is big enough to cover the region where the value of $g(x)$ is not negligible. The algorithm consists of the following four steps.

1. Determination of initial region: Determine lower and upper limits $-\infty < a_0^{(i)} < b_0^{(i)} < \infty$, such that the initial sampling region $S_0 = \prod_{i=1}^{d}[a_0^{(i)}, b_0^{(i)}] \subseteq S$, which provides an initial configuration for the support of g.

2. Discretization: With the initial region determined, we generate nd independent uniform random numbers over the initial sampling importance region S_0. We generate $x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)} \sim Unif[a_0^{(i)}, b_0^{(i)}]$ in each of the $i = 1, 2, \cdots, d$ dimensions independently. Then, we combine the d points from each dimension sequentially to form d-vectors $x_j = (x_j^{(1)}, x_j^{(2)}, \cdots, x_j^{(d)})$, $j = 1, 2, \cdots, n$, respectively. The set $S_{0,n} = (x_j, j = 1, 2, \cdots, n)$ is a discretized version of S_0.

3. Contourization: Reorder x_j, such that $g(x_k) \geqslant g(x_j)$, if $k \geqslant j$. Given an integer $c \in \mathbb{N}$, partition $S_{0,n}$ into c contours $\mathrm{E}_k = \left\{x_j : (k-1)l < j < kl, k = 1, 2, \cdots, c\right\}$, where $l = [n/c]$. Further, define a discrete distribution on the partitions $\{\mathrm{E}_k\}_{k=1}^{c}$ as

$$P_c(k) = \frac{\bar{g}_k}{\sum_{k=1}^{c} \bar{g}_k}, \qquad k = 1, 2, \cdots, c.$$

$$\bar{g}_k = \frac{1}{l} \sum_{x_j} \in \mathrm{E}_k \, g(x_j).$$

4. Sampling: First, randomly sample m subsets with replacement from $\{\mathrm{E}_k\}_{k=1}^{c}$ according to probabilities $\{P_c(k)\}_{k=1}^{c}$. Denote by m_i the number of occurrences of E_k in the m draws, where $\sum_{k=1}^{c} m_k = m$. Then for each $1 \leqslant k \leqslant c$, randomly sample m_k points with replacement from the contour E_k. In other words, each point in E_k has

equal selection probability. The set of all such points chosen from all contours forms the desired sample of size m.

As point out in Fu and Wang (2002), it is clear to see that samples constructed on the basis of this algorithm are independent identically distributed, and have a distribution which approximates $g(x)/I(g)$, when n and c are large enough. The proof is given in Fu and Wang [9].

Now, we extend the above 4-step algorithm by one additional step that determines an importance sampling region for the given distribution g.

5. Visualizing and Iteration: The algorithm of Fu and Wang (2002) is augmented in the following manner to ascertain an updated importance region. To do this, we proceed as follows:

(i) Plot histograms of the m sample points generated in Step 4 over each dimension, and calculate the maximum, minimum, and sample standard deviation, defined by $Max(x^{(i)})$, $Min(x^{(i)})$, and $Std(x^{(i)})$, $i = 1, 2, \cdots, d$, respectively, where $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \cdots, x_m^{(i)}\}$. These values will enable us to update the importance region over each dimension.

(ii) Now, set

$$a_1^{(i)} = Min(x^{(i)}) - \delta^{(i)} Std(x^{(i)}),$$

and

$$b_1^{(i)} = Max(x^{(i)}) + \delta^{(i)} Std(x^{(i)}),$$

where $i = 1, 2, \cdots, d$, and $\delta^{(i)} > 0$ is a variation adjustor, which is determined to be sufficient to cover sampling region error. This process produces a new importance sampling region $S_1 = \prod_{i=1}^{d} [a_1^{(i)}, b_1^{(i)}]$.

(iii) Return to step 1 and replace the initial importance region with the updated importance region S_1. This process is iterated until the importance sampling region for the given distribution stabilizes to a region, which we denote as $S_{it} = \prod_{i=1}^{d} [a_{it}^{(i)}, b_{it}^{(i)}]$.

At this point, we finalize an importance sampler as:

$$h(x) = \frac{\mathbb{I}_{S_{it}}(x)}{|S_{it}|}, \tag{4.1.1}$$

where $|S_{it}| = \prod_{i=1}^{d} [b_{it}^{(i)} - a_{it}^{(i)}]$ is the volume of S_{it}, and $\mathbb{I}(S_{it})$ is its indicator function.

Finally, the sample mean of $g(x)$ over the region S_{it} is an estimate of its integral over S_{it}:

$$\hat{I}_{h_{it},n} = \frac{|S_{it}|}{n} \sum_{i=1}^{n} g(X_i), \tag{4.1.2}$$

where the *i.i.d.* random sample $X_1, X_2, \cdots, X_n$ are uniformly distributed in S_{it}.

4.2 Properties of the Algorithm: One-dimensional Case

This method is applicable to many types of integrals, particularly where the value of the integral is close to 0 over the edge of its support. We begin by applying this method to one dimensional integration under two general cases and then move on to multi-dimensional integration problems. We also show how this method can reduce the *mean square error* (MSE) of the importance sampling estimator of

$$I = \int_a^b g(x)dx,$$

With the uniform importance sampler $h(x) = \frac{\mathbb{I}_{[a,b]}(x)}{b-a}$, where $\mathbb{I}$ is the indicator function. Then, the importance sampling estimator of $g(x)$ is

$$\hat{I}_{h,n} = \frac{b-a}{n}\sum_{i=1}^{n} g(X_i), \tag{4.2.1}$$

where $X_i \sim Unif[a,b]$, and its MSE is

$$\begin{aligned} MSE[\hat{I}_{h,n}] &= \mathrm{Var}(\hat{I}_{h,n}) \\ &= \frac{1}{n}\left[\int_a^b \frac{g^2(x)}{h(x)}dx - \left(\int_a^b g(x)dx\right)^2\right] \\ &= \frac{1}{n}\left[(b-a)\int_a^b g^2(x)dx - I^2(g)\right]. \end{aligned} \tag{4.2.2}$$

Case 1: We assume that $g(x) > 0$ on $[a+\xi_a, b-\xi_b]$, and is zero otherwise, where $\xi_a > 0$ and $\xi_b > 0$ are known, and such that $a + \xi_a < b - \xi_b$. Using the method of

sampling over importance region, the importance sampler assumes the form:

$$h_{it}(x) = \frac{\mathbb{I}_{[a+\xi_a, b-\xi_b]}(x)}{b - \xi_b - (a + \xi_a)}, \tag{4.2.3}$$

then, with the importance sampler h_{it} in (4.2.3), the estimated integral is

$$\hat{I}_{h_{it},n} = \frac{b - \xi_b - (a + \xi_a)}{n} \sum_{i=1}^{n} g(X_i), \qquad \text{where } X_i \sim Unif[a + \xi_a, b - \xi_b]. \tag{4.2.4}$$

We now have the following theorem,

Theorem 4.2.1. *Suppose $g(x) = 0$ in the negligible regions $[a, a + \xi_a)$ and $(b - \xi_b, b]$. The estimator produced from the method of sampling over importance region (SOIR) guarantees reduction of MSE over importance sampling method:*

$$MSE[\hat{I}_{h_{it},n}] < MSE[\hat{I}_{h,n}]. \tag{4.2.5}$$

Proof. From the algorithm introduced in Section 4.1, it is clear that we are able to reduce the sampling region from $[a, b]$ to $[a + \xi_a, b - \xi_b]$ within a certain number of iterations. We have

$$\begin{aligned} MSE[\hat{I}_{h_{it},n}] &= \mathrm{Var}(\hat{I}_{h_{it},n}) \\ &= \frac{1}{n}\left[\int_{a+\xi_a}^{b-\xi_b} \frac{g^2(x)}{h_{it}(x)} dx - \left(\int_{a+\xi_a}^{b-\xi_b} g(x)dx\right)^2\right] \\ &= \frac{1}{n}\left[\Big(b - \xi_b - (a + \xi_a)\Big) \int_{a+\xi_a}^{b-\xi_b} g^2(x)dx - I^2(g)\right]. \end{aligned} \tag{4.2.6}$$

We know that, since $g(x) = 0$, $\forall x \in [a, b] - [a + \xi_a, b - \xi_b]$, we have

$$\int_a^b g^2(x)dx = \int_{a+\xi_a}^{b-\xi_b} g^2(x)dx,$$

and also $b - \xi_b - (a + \xi_a) < b - a$. Therefore, it follows that

$$MSE[\hat{I}_{h_{it},n}] < MSE[\hat{I}_{h,n}].$$

□

For the integration problem discussed in Example 4 in Chapter 3,

$$g(x) = \begin{cases} 4\sqrt{1-x^2} & x \in [0,1], \\ 0 & x \in (1,2], \end{cases}$$

The sampling region of $g(x)$ with the methods of importance sampling (IS) and sampling over importance region (SOIR) are shown in Figure 4.1. The summarized numerical results produced from both methods are given in Table 4.1.

Table 4.1: Performance comparison between IS and SOIR for example 4

Method	IS	SOIR
Sampling region	$[0,2]$	$[0,1]$
Importance Sampler	$h(x) = \frac{1}{2}$	$h(x) = 1$
True Value	π	π
MSE	$\frac{11.4641}{n}$	$\frac{0.7968}{n}$

It is clear that $\hat{I}_{h_{it},n}$ has much smaller MSE than $\hat{I}_{h,n}$.

Case 2: Suppose we have $g(x) \approx 0$ over the interval $[a, a + \xi_a]$ and $[b - \xi_b, b]$. If we choose the uniform importance sampler $h(x) = \frac{1}{b-a}$, where $x \in [a,b]$, Equations (4.2.1) and (4.2.2) still hold for evaluating $\hat{I}_{h,n}$ and $MSE[\hat{I}_{h,n}]$.

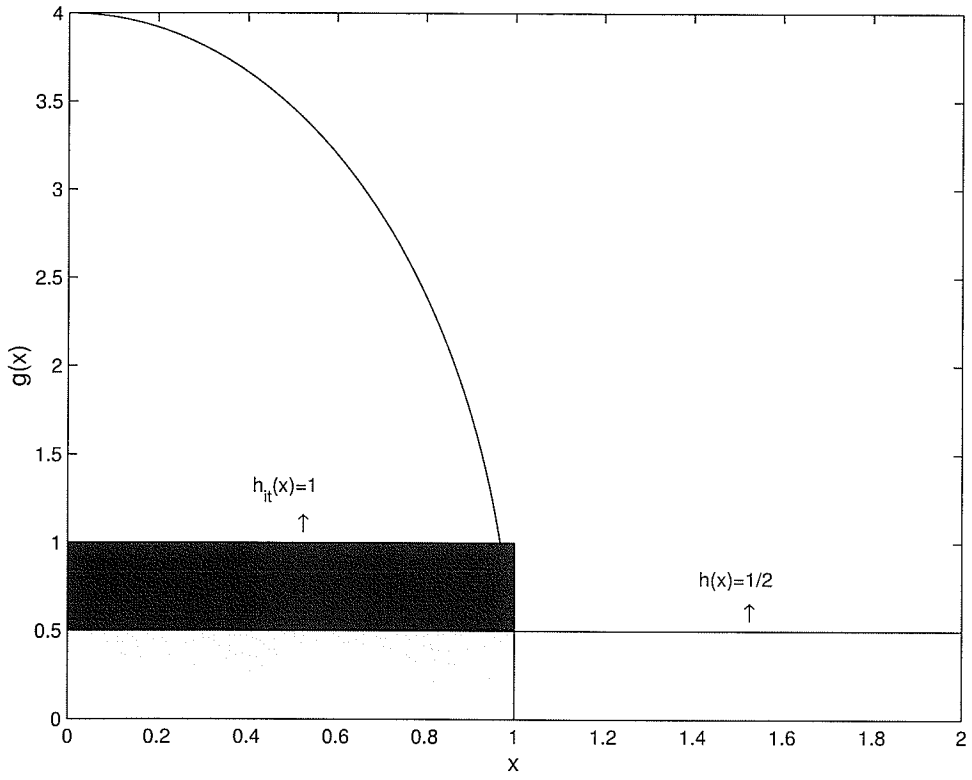

Figure 4.1: Sampling region of $g(x)$ Between IS and SOIR in example 4

The method of sampling over importance region allows us to find the important region $[a+\xi_{a_{it}}, b-\xi_{b_{it}}]$ iteratively by following the rules of choosing appropriate $\xi_{a_{it}}$ and $\xi_{b_{it}}$ specified in Theorem 4.2.2. The method of sampling over importance region yields an estimator of $I = \int_a^b g(x)dx$ as :

$$\hat{I}_{h_{it},n} = \frac{b-\xi_{b_{it}}-(a+\xi_{a_{it}})}{n}\sum_{i=1}^{n} g(X_i), \qquad \text{where } X_i \sim Unif[a+\xi_{a_{it}}, b-\xi_{b_{it}}]. \tag{4.2.7}$$

Theorem 4.2.2. *For a given sample size n, we have*

$$MSE[\hat{I}_{h_{it},n}] \leqslant MSE[\hat{I}_{h,n}] \tag{4.2.8}$$

if and only if the following two inequalities hold:

(i) The variances of $\hat{I}_{h_{it},n}$ and $\hat{I}_{h,n}$ satisfy

$$Var[\hat{I}_{h_{it},n}] \leqslant Var[\hat{I}_{h,n}] \tag{4.2.9}$$

(ii) $\xi_{a_{it}}$ and $\xi_{b_{it}}$ satisfy

$$\left(\int_a^{a+a_{\xi_{it}}} g(x)dx + \int_{b-b_{\xi_{it}}}^{b} g(x)dx\right)^2 \leqslant Var[\hat{I}_{h,n}] - Var[\hat{I}_{h_{it},n}]. \tag{4.2.10}$$

Proof. Using the method of sampling over importance regions, the *mean square error* of the estimator $\hat{I}_{h_{it},n}$ (given in (4.2.7)) is:

$$\begin{aligned} MSE[\hat{I}_{h_{it},n}] &= \left[I - \mathrm{E}(\hat{I}_{h_{it},n})\right]^2 + \mathrm{Var}(\hat{I}_{h_{it},n}) \\ &= \left(\int_a^b g(x)dx - \int_{a+a_{\xi_{it}}}^{b-b_{\xi_{it}}} g(x)dx\right)^2 + \mathrm{Var}(\hat{I}_{h_{it},n}) \qquad (4.2.11) \\ &= \left(\int_a^{a+a_{\xi_{it}}} g(x)dx + \int_{b-b_{\xi_{it}}}^{b} g(x)dx\right)^2 + \mathrm{Var}(\hat{I}_{h_{it},n}). \end{aligned}$$

Also we have (4.2.2) as following

$$MSE[\hat{I}_{h,n}] = \mathrm{Var}(\hat{I}_{h,n}).$$

Therefore, the theorem follows from (4.2.2) and (4.2.11). □

Let's see an example to demonstrate the situation in Case 2.

Example 5. Suppose we want to integrate $g(x) = x^{-5}$ over the interval $x \in [1, 10]$. Of course $I = \frac{1}{4}(1 - 10^{-4})$. The uniform importance sampler will be

$$h(x) = \frac{1}{9}, \quad \text{where } x \in [1, 10],$$

which will give us an unbiased estimator

$$\hat{I}_{h,n} = \frac{9}{n}\sum_{i=1}^{n} g(X_i) \quad X_i \sim Unif[1, 10],$$

with variance

$$\mathrm{Var}[\hat{I}_{h,n}] = \frac{1}{n}\left[\int_1^{10} 9\, g^2(x)dx - I^2\right] = \frac{15}{16n}.$$

Now, let's consider the method of sampling over importance region. From Figure 4.2, we clearly see that $g(x) \approx 0 \ \ \forall x \in [4, 10]$. Indeed, the algorithm of sampling over importance region allows us to pick appropriate $\xi_{a_{it}}$ and $\xi_{b_{it}}$ by iteration, the corresponding estimated integral has smaller MSE than that of $\hat{I}_{h,n}$.

Suppose the SOIR algorithm determines on $\xi_{a_{it}} = 0$, and $\xi_{b_{it}} = 6$. This will lead to an importance sampling region $[1, 4]$, and our updated uniform importance sampler

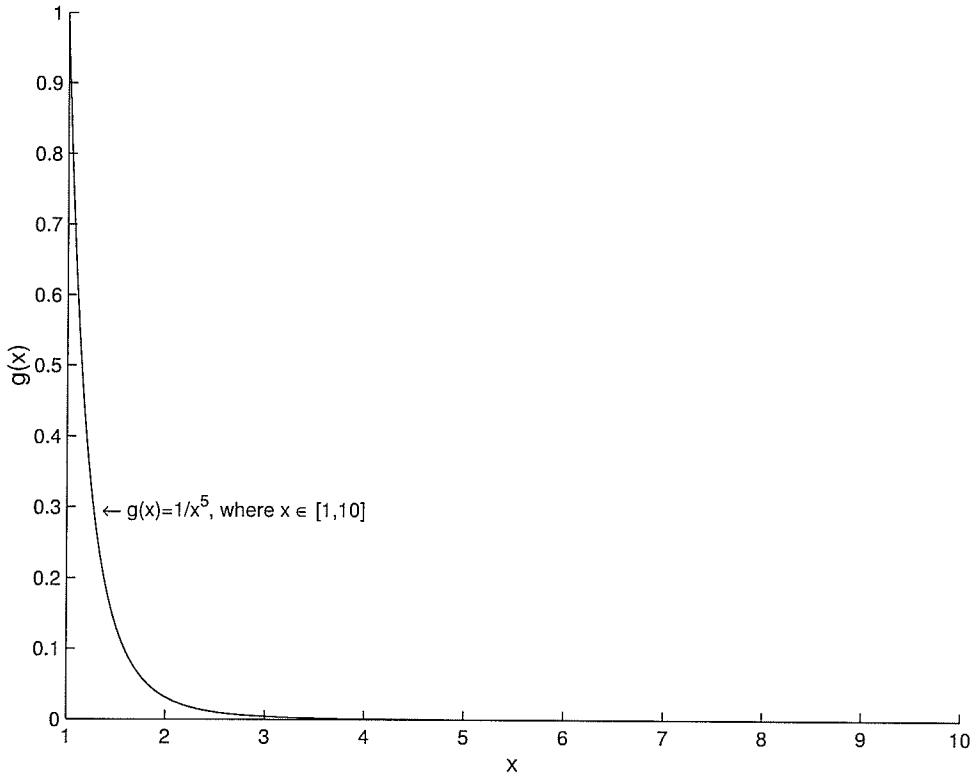

Figure 4.2: Distribution of g(x) of example 5

is

$$h_{it}(x) = \frac{1}{3} \quad x \in [1, 4],$$

which gives an unbiased estimator for $I_{it} = \int_1^4 x^{-5} dx$.

$$\hat{I}_{h_{it},n} = \frac{3}{n} \sum_{i=1}^{n} g(x_i) \quad \text{where } x \in h_{it}(x).$$

We also have

$$\text{Var}[\hat{I}_{h_{it},n}] = \frac{1}{n} \left[\int_1^4 3\, g^2(x) dx - I_{it}^2 \right] = \frac{13}{48n}.$$

Obviously, $\text{Var}[\hat{I}_{h_{it},n}] < \text{Var}[\hat{I}_{h,n}]$, which satisfies Condition 1 of Theorem 4.2.2. Furthermore,

$$\begin{aligned} \left(\int_a^{a+a_{\xi_{it}}} g(x) dx + \int_{b-b_{\xi_{it}}}^{b} g(x) dx \right)^2 &= \frac{1}{16} \left[(4^{-4} - 10^{-4}) \right]^2 \\ &\leqslant \text{Var}[\hat{I}_{h,n}] - \text{Var}[\hat{I}_{h_{it},n}] \\ &= \frac{15}{16n} - \frac{13}{48n} \\ &= \frac{32}{48n}, \end{aligned} \tag{4.2.12}$$

which requires sample size $n \leqslant 46016$ for the specified $\xi_{a_{it}}$ and $\xi_{b_{it}}$. From Theorem 4.2.2, we have

$$MSE[\hat{I}_{h_{it},n}] \leqslant MSE[\hat{I}_{h,n}].$$

Therefore, we obtained a better estimator by sampling over the importance region in terms of smaller MSE. The numerical results of comparison between importance sampling method (IS) and sampling over importance region method (SOIR) for Example 5 are summarized in Table 4.2.2.

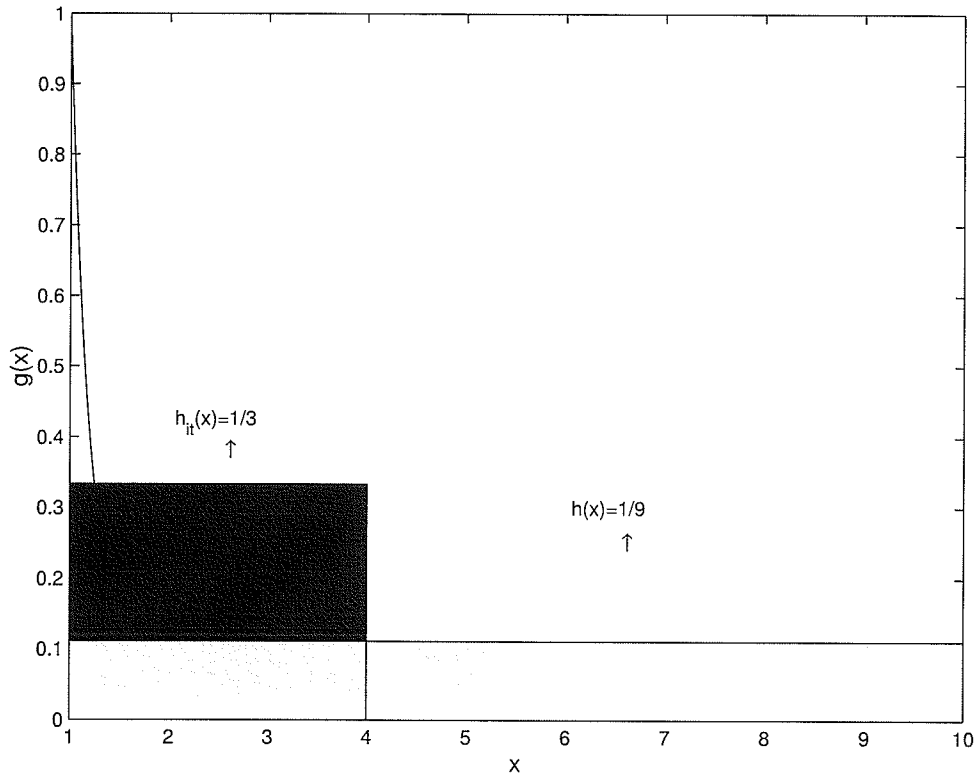

Figure 4.3: Sampling region of $g(x)$ with IS and SOIR in example 5

Table 4.2: Performance comparison between IS and SOIR in example 5

Method	IS	SOIR
Sampling region	$[1, 10]$	$[1, 4]$
Importance Sampler	$h(x) = \frac{1}{9}$	$h(x) = \frac{1}{3}$
MSE	$\frac{15}{16n}$	$\approx \frac{13}{48n} + 10^{-5}$

4.3 Properties of the Algorithm: Multi-Dimensional Case

In this Section, we extend the idea described in Section 4.2 to multi-dimensional problems. Suppose we want to estimate $I = \int_S g(x)dx$, where $S = \prod_{i=1}^{d}[a^{(i)}, b^{(i)}]$. If an importance sampler $h(x)$ is chosen as

$$h(x) = \frac{\mathbb{I}_S(x)}{|S|}, \tag{4.3.1}$$

where $|S| = \prod_{i=1}^{d}[a^{(i)}, b^{(i)}]$ is the volume of S, and $\mathbb{I}_S(x)$ is its indicator function. Then, the importance sampling estimator of $I(g)$ is

$$\hat{I}_{h,n} = \frac{|S|}{n}\sum_{i=1}^{n} g(X_i), \tag{4.3.2}$$

where the *i.i.d.* random sample $X_1, X_2, \cdots, X_n$ is uniformly distributed in S. Furthermore, its MSE is

$$\begin{aligned} \text{MSE}[\hat{I}_{h,n}] &= \text{V}ar(\hat{I}_{h,n}) \\ &= \frac{1}{n}\left[\int_S \frac{g^2(x)}{h(x)}dx - \left(\int_S g(x)dx\right)^2\right] \\ &= \frac{1}{n}\left[|S|\int g^2(x)dx - I^2(g)\right] \end{aligned} \tag{4.3.3}$$

Case 1: We consider $g(x) = 0$ over known intervals $[a^{(i)}, a^{(i)} + \xi_a^{(i)})$ and $(b^{(i)} - \xi_b^{(i)}, b^{(i)}]$, where $\xi_a^{(i)} > 0$, $\xi_b^{(i)} > 0$, and $i = 1, 2, \cdots, d$.

Now, we define $S_{it} = \prod_{i=1}^{d}[a^{(i)} + \xi_a^{(i)}, b^{(i)} - \xi_b^{(i)}]$ to be the importance region at the last iteration, from which we form the importance sampler:

$$h(x) = \frac{\mathbb{I}_{S_{it}}(x)}{|S_{it}|}, \tag{4.3.4}$$

where $|S_{it}|$ is the volume of S_{it}, and $\mathbb{I}_{S_{it}}(x)$ is its indicator function.

Theorem 4.3.1. *Suppose $g(x) = 0$ in the negligible regions $\prod_{i=1}^{d}[a^{(i)}, a^{(i)} + \xi_a^{(i)}]$ and $\prod_{i=1}^{d}(b^{(i)} - \xi_b^{(i)}, b^{(i)}]$. The estimator produced from the method of sampling over importance region is guaranteed to reduce the MSE over that from the importance sampling method; i.e.,*

$$MSE[\hat{I}_{h_{it},n}] < MSE[\hat{I}_{h,n}]. \tag{4.3.5}$$

Proof. From the algorithm introduced in Section 4.1, it is clear that we are able to reduce the sampling region from S to S_{it} within a finite number of iterations. Therefore, with the importance sampler h_{it} in (4.3.4), we have

$$\hat{I}_{h_{it},n} = \frac{|S_{it}|}{n} \sum_{i=1}^{n} g(X_i), \tag{4.3.6}$$

where $\{X_1, X_2, \cdots, X_n\} \in S_{it}$ are $i.i.d.$ uniformly distributed random variables. Since $|S_{it}| = \prod_{i=1}^{d} [b^{(i)} - \xi_b^{(i)} - (a^{(i)} + \xi_a^{(i)})]$, and

$$\mathrm{E}(\hat{I}_{h_{it},n}) = \int_{S_{it}} g(x)dx = I.$$

It follows that

$$\begin{aligned} MSE[\hat{I}_{h_{it},n}] &= \mathrm{Var}(\hat{I}_{h_{it},n}) \\ &= \frac{1}{n}\left[\int_{S_{it}} \frac{g^2(x)}{h_{it}(x)}dx - \left(\int_{S_{it}} g(x)dx\right)^2\right] \\ &= \frac{1}{n}\left[|S_{it}| \int g^2(x)dx - I^2(g)\right]. \end{aligned} \tag{4.3.7}$$

We also know that since $g(x) = 0 \ \forall \ x \in S - S_{it}$,

$$\int_S g^2(x)dx = \int_{S_{it}} g^2(x)dx$$

and also $\xi_a^{(i)} > 0$ and $\xi_b^{(i)} > 0$, where $i = 1, 2, \cdots, d$, implies $|S_{it}| < |S|$ Therefore, from (4.3.3) and (4.3.7), it is obvious that

$$MSE[\hat{I}_{h_{it},n}] < MSE[\hat{I}_{h,n}].$$

□

Under above situation, our integration approximation method by sampling over importance region establishes an unbiased estimator, and guarantees the MSE of this estimator is smaller than the estimator obtained by using importance sampling method over the entire region S.

Case 2: Suppose we want to estimate $I = \int_S g(x)dx$, where $S = \prod_{i=1}^{d}[a^{(i)}, b^{(i)}], i = 1, 2, \cdots, d$. Moreover, $g(x) \approx 0$ over the intervals $[a^{(i)}, a^{(i)} + \xi_a^{(i)}]$ and $[b^{(i)} - \xi_b^{(i)}, b^{(i)}]$. If we choose the uniform importance sampler $h(x)$ as in (4.3.1), Equations (4.3.2) and(4.3.3) hold for evaluating $\hat{I}_{h,n}$ and $MSE[\hat{I}_{h,n}]$.

Integration by the method of sampling over importance region allows us to find the important region $S_{it} = \prod_{i=1}^{d}[a^{(i)} + \xi_{a_{it}}^{(i)}, b^{(i)} - \xi_{b_{it}}^{(i)}]$ by iteratively following the rule of choosing appropriate $\bar{\xi}_{a_{it}}$ and $\bar{\xi}_{b_{it}}$ so that we can form an importance sampler:

$$h_{it}(x) = \frac{\mathbb{I}_{S_{it}}(x)}{|S_{it}|}. \tag{4.3.8}$$

For a given sample size n, we define $\hat{I}_{h_{it},n}$ as the estimator of $I = \int_S g(x)dx$.

$$\hat{I}_{h,n} = \frac{|S_{it}|}{n}\sum_{i=1}^{n} g(X_i), \tag{4.3.9}$$

where $X_i \sim Unif(S_{it})$. Then, we have the following result:

Theorem 4.3.2. *Suppose we are evaluating $\int_S g(x)dx$. For a given sample size n, $MSE[\hat{I}_{h_{it},n}]$ and $MSE[\hat{I}_{h,n}]$ are the mean square error of $\hat{I}_{h_{it},n}$ and $\hat{I}_{h,n}$ respectively. Then,*

$$MSE[\hat{I}_{h_{it},n}] \leqslant MSE[\hat{I}_{h,n}] \tag{4.3.10}$$

if and only if the following two inequalities hold:

(i) The variance of $\hat{I}_{h_{it},n}$ *and* $\hat{I}_{h,n}$ *satisfy*

$$Var[\hat{I}_{h_{it},n}] \leqslant Var[\hat{I}_{h,n}] \tag{4.3.11}$$

(ii) The bias of $\hat{I}_{h_{it},n}$ *satisfies*

$$\Big(\int_{S-S_{it}} g(x)dx\Big)^2 \leqslant Var[\hat{I}_{h,n}] - Var[\hat{I}_{h_{it},n}] \tag{4.3.12}$$

where S *and* S_{it} *are the support of* $g(x)$ *before and after applying the method of SOIR, respectively.*

Proof. The proof is an the expansion of the proof of Theorem 4.2.2. □

4.4 Sampling Over Importance Regions with Unbounded Support

If either the upper limit or lower limit of the support of g is unbounded over a particular dimension, we introduce the following two general methods to convert the unbounded sampling region to a bounded one, and set the proper initial sampling region. For simplicity, we only assume the upper limit $b^{(i)}$ is unbounded. It is readily extended to the situation where both limits are unbounded.

(i) Method of Transformation: Consider integrating function $g(x_1, x_2, \cdots, x_d)$ on $S \subseteq \mathbb{R}^d$. Often, there exists a one-to-one transformation from $S \to D \subseteq \mathbb{R}^d$

$$y_i = u_i(x_1, x_2, \cdots, x_d),\ i = 1, 2, \cdots, d,$$

with inverse transformation

$$x_j = w_i(y_1, y_2, \cdots, y_d),\ j = 1, 2, \cdots, d.$$

If the Jacobian $J = \left(\frac{\partial x_i}{\partial y_i}\right)_{d\times d}$ is continuous and nonzero over D, then,

$$\int_S g(x)dx = \int_D g[w_1(y_1, \cdots, y_d), \cdots, w_k(y_1, \cdots, y_d)]|J|dy. \tag{4.4.1}$$

This method allows us to map the set of all points $x = (x_1, x_2, \cdots, x_d)$ into $y = (y_1, y_2, \cdots, y_d)$ from S to D, where D is bounded. Generally speaking, we can transform any monotone function with unbounded support to a function with bounded support. Consequently, the initial sampling region will be set as D, which has bounded lower and upper limit.

Once we carefully transform an integral with unbounded support to a problem with bounded support, the algorithm of sampling over importance region described in Section 4.2 can be applied.

Let's take isotropic function as an example. A class of isotropic test functions of Capstick and Keister (1996) takes the following form with unbounded support $\mathbb{R}^d$,

$$I = \int_{\mathbb{R}^d} h(\| x \|) \exp^{-\|x\|^2} dx, \tag{4.4.2}$$

where $\| x \| = \sqrt{x'x}$. It can be transformed to the following integral over bounded support $[0,1]^d$:

$$I = \pi^{d/2} \int_{[0,1]^d} h\left(\sqrt{\frac{1}{2}\sum_{j=1}^{d}[\Phi^{-1}(x^j)]^2}\right) dx, \tag{4.4.3}$$

where $\Phi(x)$ is the standard normal distribution function.

Although the method of transformation can solve the problems of unbounded regions theoretically, it has limitations when integrating over high dimensions. Sometimes, it can make the integration problem even more complicated after transformation. In such cases, we introduce another simpler approximation method.

(ii) Method of Approximation: By the properties of the integral and common sense, we can find a sufficient large number $M^{(i)}$ which satisfies

$$\left| \int_{\prod[a^{(i)}, M^{(i)}]} g(x)dx - \int_{\prod[a^{(i)}, \infty]} g(x)dx \right| < \varepsilon, \tag{4.4.4}$$

where $\varepsilon > 0$ is any arbitrarily small number. For example, if $g(x)$ is a multi-normal distribution, then it is pretty safe to set $M^{(i)} = \mu^{(i)} + 6\sigma^{(i)}$. Therefore, we can denote $b_0^{(i)} = M^{(i)}$ as the initial upper limits. By doing so, we have a small bias $B = \int_{\prod[a^{(i)}, M^{(i)}]} g(x)dx - \int_{\prod[a^{(i)}, \infty]} g(x)dx$. Therefore, similar to the Case 2 of Section 4.3, the values of $M^{(i)}$ should be carefully chosen to be large enough such that the squared bias B^2 is restricted to a small range.

Chapter 5

Benchmark Examples and Applications

In this Chapter, we demonstrate some benchmark numerical integration examples by using the method of sampling over importance region (SOIR). We compare the numerical results with those from using existing integration methods. We also apply the method of SOIR to widely used mixture functions, isotropic functions and Genz test functions, and show the method of sampling over importance region can be used as a general tool for numerical integration.

5.1 Examples

We have applied this method to a number of examples, including some benchmark examples which have frequently appeared in the recent literature.

Example 6. Evans and Swartz (1995) considered an integral of density function $g(x)$ of the six dimensional normal distribution $N_6(0, \Sigma)$ over the region $[0, \infty)^6$, where

$\Sigma^{1/2} = \text{diag}(0, 1, 2, 3, 4, 5) + ee'$ and $e = (1, 1, 1, 1, 1, 1)'$. As pointed out by Evans and Swartz (1995), an accurate numerical evaluation of this integral is extremely difficult. The value of $g(x)$ is very small for $x\varepsilon[0, \infty)^6$. In fact, the value of this integral is approximately $\int_{[0,\infty)^6} g(x)dx \approx 1/60,000$, which means that in about every $60,000$ generated observations we expect to have one observation falling in the region $[0, \infty)^6$. This makes it very difficult to use any importance sampling method to find a good estimate. However, our method provides a solution for this problem.

We now demonstrate detailed steps in according to the algorithm described in Section 4.1.

Step 1: We determine the initial region. We notice that this is a problem with unbounded support region. Therefore, using the method stated in Section 4.4, we carefully choose a upper limit M, which is sufficiently large, say 10. Then, the initial sampling region is $S_0 = [0, 10]^6$.

Step 2: We discretize S_0 by $S_{0,n}$. In Matlab, we first generate a total of 6×10^6 independent uniform random numbers with $n_0 = 10^6$ numbers on each dimension. Then, we choose a point from each dimension sequentially and combine them to form 6-vectors $x_j = (x_j^{(1)}, x_j^{(2)}, \cdots, x_j^{(6)})$, $j = 1, 2, \cdots, n_0$. As a result, we have the set $S_{0,n} = \{x_j, j = 1, 2, \cdots, n_0\}$ as a discretized version of S_0.

Step 3: We contourize the discretized support S_0, n. First, we order $x_j = (x_j^{(i)})$, where $i = 1, 2, \cdots, 6$ and $j = 1, 2, \cdots, n_0$, respectively. Those points x_j which make the value $g(x) \approx 0$ are ignored. Then, we will have the actual number of base points, denoted as n_1 ($n_1 \leqslant n_0$ of course). Next, we partition $S_{0,n}$ into $c = n_1/l$ contours, where $l = 30$ is predetermined parameter to be the number of points in each contour. With the parameters n_0 and l initiated, Matlab program is set up to realize the procedures and produces $n_1 = 2324$ base points and $c = 77$ contours.

Further, we defined a discrete distribution on the partitions $\{E_k\}_{k=1}^{c}$ as

$$P_c(k) = \frac{\bar{g}_k}{\sum_{k=1}^{c} \bar{g}_k}, \qquad k = 1, 2, \cdots, 77.$$

$$\bar{g}_k = \frac{1}{30} \sum_{x_j} \in \mathrm{E}_k g(x_j) \qquad \text{where } j = 1, 2, \cdots, n_1.$$

Step 4: Sample $m = 6000$ subsets with replacement from the contourized support $\{E_k\}_{k=1}^{77}$ according to the probabilities $\{P_c(k)\}_{k=1}^{77}$, m_k is the number of occurrences of E_k in the m draws, where $\sum_{k=1}^{77} m_k = m = 6000$.

Step 5: We examine the selected marginal histograms over each dimension as in Figure 5.1. From these histograms, we can clearly see that because of the variance-covariance structure of this distribution, the density function highly concentrates around the origin and is almost zero anywhere else. Therefore, it is safe to reduce the upper limit to be 3 for the second iteration in order to cover the importance sampling region. This produces a new importance sampling region of $g(x)$ as $S_1 = [0, 3]^6$.

Next, we return to Step 1, and replace S_0 with S_1. The algorithm of sampling over

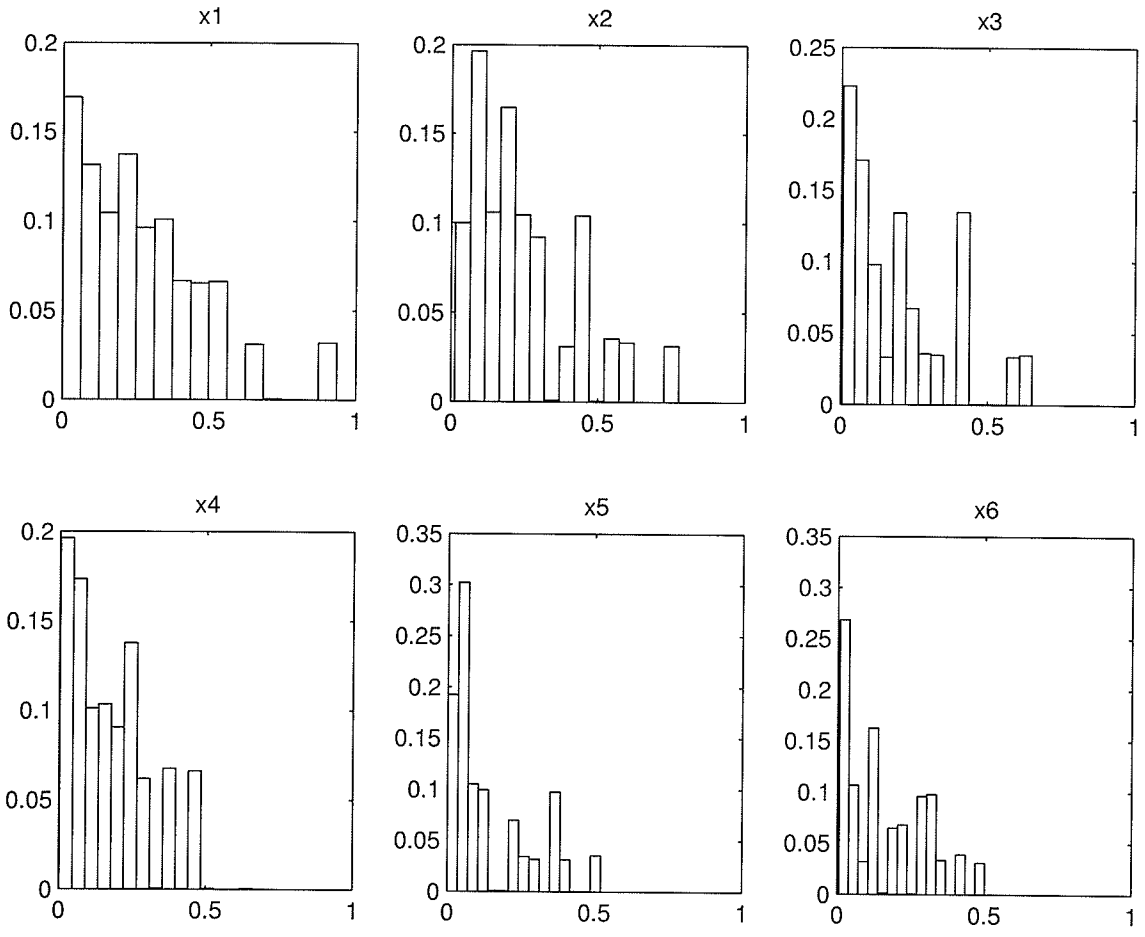

Figure 5.1: Histogram of marginal distribution of example 6 over initial importance sampling region

importance region described in Section 4.3 allows us to find the important region by choosing appropriate lower and upper bounds for sampling over each dimension of $g(x)$ after a few iterations. As a result, an importance region $S = [0,1]\times[0,1]\times[0,1]\times[0,0.5]\times[0,0.5]\times[0,0.5]$ is finalized after the $6th$ iteration. We then sample 10^6 uniform

observations from S to yield a sample mean for $g(x)$ as 1.663×10^{-5}. This value is very close to the result obtained by Evans and Swartz (1995)—1.66625×10^{-5}, with much more sophisticated procedures. Figure 5.2 shows the marginal histograms over each dimension of $g(x)$ after the last iteration. As a remark, the determination of the

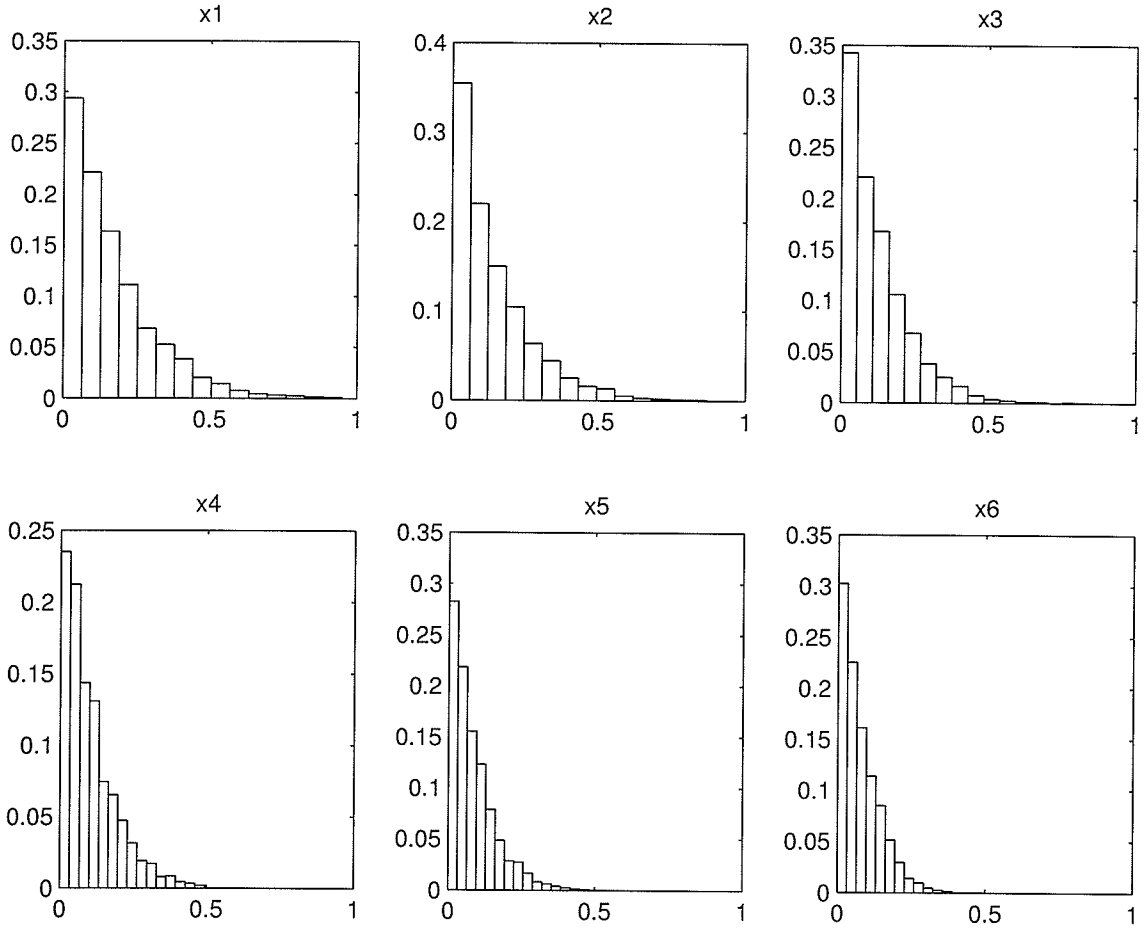

Figure 5.2: Histogram of marginal distribution of example 6 over improved importance sampling region

initial parameters of n_0 and l is based on the property of the integrand, the number of

dimensions of the integral, and the capacity of the computer, etc. Intuitively, we need n_0 to be large enough, and l to be small enough to ensure the quality of the estimate. However, the sizes of these parameters are normally limited by the constraints above.

Example 7. We continue to work on Example 3 in Chapter 3, where

$$g(x) = 0.25g_1(x) + 0.75g_2(x) \tag{5.1.1}$$

$x = (x_1, x_2)' \in \mathbb{R}^2$, and g_1, g_2 are the density functions corresponding to the $\mathrm{N}(\mu_1, \Sigma_1)$ and $\mathrm{N}(\mu_2, \Sigma_2)$ distributions, where

$$\mu_1 = \begin{pmatrix} 0.0 \\ \\ 0.0 \end{pmatrix}, \qquad \mu_2 = \begin{pmatrix} 2.1 \\ \\ 2.1 \end{pmatrix},$$

and

$$\Sigma_1 = \begin{pmatrix} 1.0 & 0.8 \\ \\ 0.8 & 1.0 \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} 1.0 & 0.0 \\ \\ 0.0 & 1.0 \end{pmatrix}.$$

respectively. We apply the procedures of sampling over importance region described in Section 4.1 to this problem. We use $n_0 = 10^6$ (number of generated random samples), and $l = 50$ (number of points in each contour). Figure 5.3 shows the improved marginal histograms of X_1 and X_2 after 6 iterations, which results in the importance sampling region $S = [-2, 6] \times [0, 6]$. The estimation results are given in the Table 5.1 for the comparison between importance sampling (IS), adaptive importance sampling (AIS) and sampling over important region (SOIR). $\hat{\mu}_1$, $\hat{\mu}_2$ are the weighted

average of 50 estimates of the 1st and 2nd elements of $\hat{\mu}$, respectively. We clearly see the better performance of SOIR over both IS and AIS.

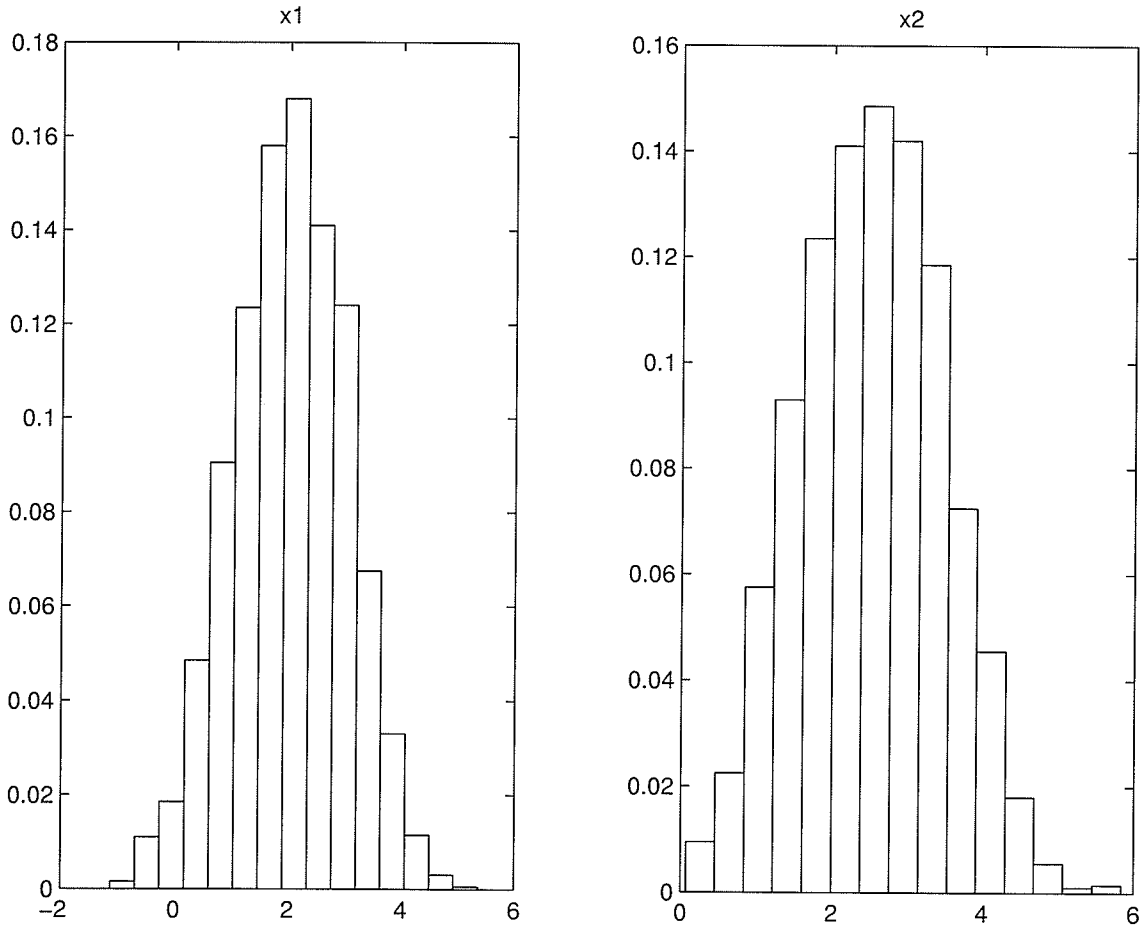

Figure 5.3: Histogram of marginal distribution of example 7 over improved importance sampling region

Now, consider some other high dimensional integration problems that have been evaluated by the adaptive importance sampling method (Karaivanova and Dimov, 1998). The exact values of integrals are presented to enable the comparison.

Table 5.1: Comparison of IS, AIS and SOIR in example 7

Parameter	IS	AIS	SOIR	True Values
n	57,600	33,000	30,000	
$\hat{\mu}_1$	1.5934	1.5837	1.5803	1.5750
$\hat{\mu}_2$	1.5986	1.5953	1.5733	1.5750
CPU Time	295.0	169.3	96.125	

Example 8. Suppose we wish to evaluate $\int_{[0,1]^{30}} g(x)\,dx$, where

$$g(x) = \frac{4x_1x_3^2\,\exp(2x_1x_3)}{(1+x_2+x_4)^2}\,\exp(x_5+\cdots+x_{20})\,x_{21}x_{22}\cdots x_{30} \tag{5.1.2}$$

The exact value of this integral is

$$I(g) = \int_{[0,1]^{30}} g(x)\,dx = 3.244$$

We will briefly describe how to apply the method of sampling over importance region to this particular problem, and compare the results with those from using AIS.

Step 1: We first determine the initial support region of $g(x)$. We notice that this is a problem with bounded support region, which is the case 2 problem as described in Section 4.3. Therefore, we set $S_0 = [0,1]^{30}$ to be initial sampling region.

Step 2: We discretize the initial support S_0 to be $S_{0,n}$. In Matlab, we first generate a total of 30×10^6 independent uniform random numbers with $n_0 = 10^6$ numbers on each dimension. Then, we choose a point from each dimension sequentially and

combine them to form 30-vectors $x_j = (x_j^{(1)}, x_j^{(2)}, \cdots, x_j^{(30)})$, $j = 1, 2, \cdots, n_0$. As a result, we have the set $S_{0,n} = \{x_j, j = 1, 2, \cdots, n_0\}$ as a discretized version of S_0.

Step 3: We contourize the discretized support of $g(x)$. First, we order $\{x_j, j = 1, 2, \cdots, n_0\}$. Those points x_j which make the value $g(x) \approx 0$ are ignored. Then, we will have the actual number of base points, denoted as n_1. Next, we partition $S_{0,n}$ into $c = n_1/l$ contours, where $l = 30$ is predetermined parameter to be the number of points in each contour. With the parameters n_0 and l initiated, Matlab program is set up to realize the procedures and produces $n_1 = 10^6$ base points and $c = 33,333$ contours.

Further, we defined a discrete distribution on the partitions $\{E_k\}_{k=1}^c$ as

$$P_c(k) = \frac{\bar{g}_k}{\sum_{k=1}^c \bar{g}_k}, \qquad k = 1, 2, \cdots, c.$$

$$\bar{g}_k = \frac{1}{30} \sum_{x_j} \in \mathrm{E}_k g(x_j) \qquad \text{where } j = 1, 2, \cdots, n_1.$$

Step 4: Sample $m = 8000$ subsets with replacement from the contourized support $\{E_k\}_{k=1}^c$ according to the probabilities $\{P_c(k)\}_{k=1}^c$, m_k is the number of occurrences of E_k in the m draws, where $\sum_{k=1}^c m_k = m = 8000$.

Step 5: We examine the selected marginal histograms over each dimension. The algorithm of sampling over importance region described in Section 4.3 allows us to find the important region by reducing lower and upper bounds to appropriate levels

after the $4th$ iteration. Here, we show the improved marginal histograms generated through the SOIR technique for Example 8 in Figures 5.4 and 5.5. It can be seen that the sampling region has been reduced from $[0, 1]$ to $[0.2, 1]$ for the third coordinate of x.

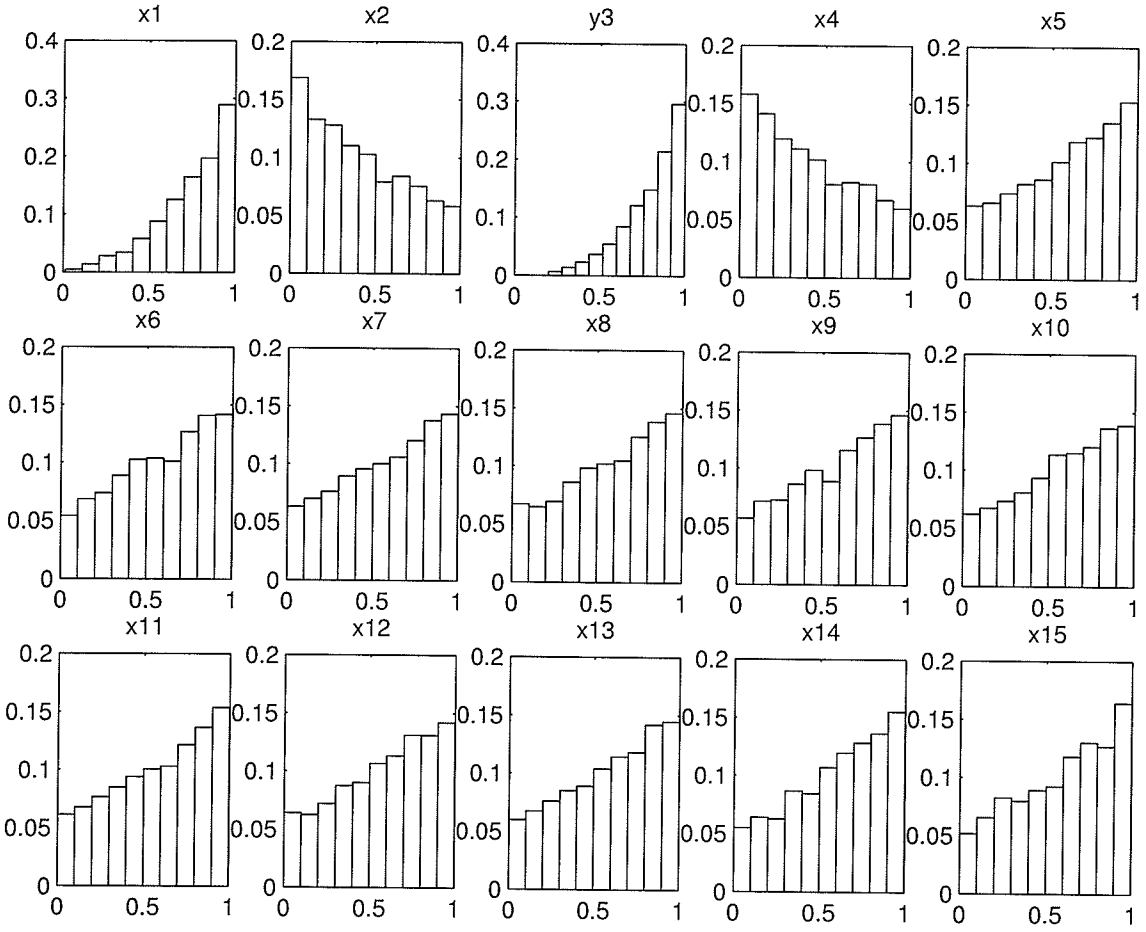

Figure 5.4: Histogram of marginal distribution of example 8 over importance region - part 1

Let us apply the method of SOIR to another two integration problems. (Karaivanova and Dimov, 1998)

Example 9.

$$g(x) = \frac{4x_1x_3^2 \exp(2x_1x_3)}{(1+x_2+x_4)^2}; \tag{5.1.3}$$

and

Example 10.

$$g(x) = \frac{4x_1x_3^2\exp(2x_1x_3)}{(1+x_2+x_4)^2}\exp(x_5+\cdots+x_{20})\,x_{21}x_{22}\cdots x_{25}. \tag{5.1.4}$$

The exact values of $I(g)$ in Example 9 and Example 10 are 0.5753 and 103.8, respectively. The procedure for evaluating the integrals with the method of SOIR in Example 9 and Example 10 is similar to that of in Example 8. I will only present the numerical results of these two examples.

We compare the numerical results of examples 8, 9, and 10 produced by using SOIR with those produced by using AIS in Table 5.2, such as the information about the dimension of the integral, number of random points, the exact solution, CPU-Time (s), relative error. From it, we conclude that the implementation of the sampling over important region (SOIR) yields a relatively accurate estimate to the target with relatively small error and less computing time.

Table 5.2: Numerical Comparison of AIS and SOIR of Example 8, 9, and 10

Dimension	$d=4$		$d=25$		$d=30$	
Sampling method	AIS	SOIR	AIS	SOIR	AIS	SOIR
n	2×10^4	2×10^4	1×10^5	1×10^5	1×10^5	1×10^5
Exact Solution	0.5753	0.5753	103.8	103.8	3.244	3.244
Calc.Solution	0.5763	0.5754	107.66	104.22	3.365	3.252
CPU Time (s)	0.184	0.203	3.338	0.7456	4.07	0.858
Relative Error	0.0017	0.0002	0.036	0.004	0.037	0.0025

5.2 Applications

The method of SOIR is applicable to many moderately high dimension integration problems. To illustrate this, we apply it to some well known test functions, such as mixture functions, isotropic functions, Genz test functions.

Mixture of Beta functions: As mentioned in previous chapters, traditional importance sampling technique has a major drawback. If $h(x)$ decreases toward 0 faster than $g^2(x)h^2(x)$ as x moves away from its modes, we can experience that $\mathrm{Var}(\hat{I}_n) = \infty$. The irony is that this infinite variance is due to a region of S that is unimportant in ordinary Monte Carlo sampling. To illustrate this point, we present the following example with the univariate beta density function

$$B(x, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \qquad 0 < x < 1,$$

where $a > 0$ and $b > 0$ are parameters, and $\Gamma(z)$ is the gamma function.

Example 11. Let $S = [0,1]^5$, the integrand is

$$g(x) = 0.9 \prod_{j=1}^{5} B(x_j, 20, 20) + 0.1 \prod_{j=1}^{5} B(x_j, 2, 2), \tag{5.2.1}$$

where $x = (x_1, x_2, \cdots, x_5) \in S$. If we evaluate this integral with the method of importance sampling (IS), we would probably choose the importance sampling density as

$$h(x) = \prod_{j=1}^{5} B(x_j, 20, 20). \tag{5.2.2}$$

This density is very nearly proportional to $g(x)$, so one might expected a good estimate. In fact, $\hat{I}_{h,n} \approx 1$. However, it can be showed that $\mathrm{Var}[\hat{I}_{h,n}] \approx \infty$.

By presetting $n_0 = 5 \times 10^5$ and $l = 30$, we apply the method of sampling over importance region described in Section 4.1 to this integration problem. The results appear to be much more acceptable. To visualize the problem, we display the marginal histograms over each of 5 dimensions after the $5th$ iteration in Figure 5.6. It resulted a improved importance support region of $g(x)$ as $S_{it} = [0.1, 0.9]^5$.

Next, with finalized importance sampling region of $g(x)$, we calculate the estimate of $I(g)$ to be $\hat{I}_{h_4,n} \doteq 0.9812$, which is the weighted averages of 100 estimators calculated over the importance region S_{it}, and $\mathrm{Var}[\hat{I}_{h_5,n}] = 0.0379$.

Isotropic functions: A class of isotropic test functions due to Capstick and Keister (1996) have been used as numerical integration test functions. Papageorgiou and Traub (1997) report good results for quasi-Monte Carlo on these functions, and Papageorgiou (2001) introduces a radial discrepancy measure for them. Novak, Ritter, Schmitt, and Steinbauer (1997) apply an interpolatory rule to some such functions. The isotropic integrals takes the form

$$\int_{\mathbb{R}^d} h(\| x \|) \exp^{-\|x\|^2} dx = \pi^{d/2} \int_{[0,1]^d} h\left(\sqrt{\frac{1}{2}\sum_{j=1}^{d}[\Phi^{-1}(x^j)]^2}\right) dx, \qquad (5.2.3)$$

where $\| x \| = \sqrt{x'x}$ and $\Phi(x)$ is the standard normal distribution function.

Capstick and Keister (1996) consider the isotropic integral with $h(z) = \cos(z)$. In

this case, the integral (5.2.3) can be expressed in terms of

$$\begin{aligned} I_d &= \pi^{-d/2} \int_{\mathbb{R}^d} \cos(\| x \|) e^{-\|x\|^2} dx \\ &= \int_{[0,1]^d} \cos\left(\sqrt{\frac{1}{2} \sum_{j=1}^{d} [\Phi^{-1}(x^j)]^2} \right) dx. \end{aligned} \tag{5.2.4}$$

The exact value of this special integral can be closely approximated by the method of transformation. Here, we apply the method of SOIR to this particular integration problem with $d = 20$. By setting $n_0 = 1 \times 10^6$ and $l = 30$, we draw the marginal histograms over importance region for each of 20 dimensions in Figure 5.7 and Figure 5.8. We also present the estimated results by applying SOIR with comparison to those yielded from its transform function for the case $d = 20$ in Table 5.3. It shows that the method of SOIR produces acceptable accurate estimation for this moderately high dimensional integration problem.

Table 5.3: Comparison of Transformation method and SOIR in Capstick and Keister (1996) example

Parameter	Transformation Method	SOIR
d	20	20
$\hat{I}$	-0.8847	-0.8451
Var$[\hat{I}]$	0.0483	0.00658

Genz test functions: We now test method of SOIR with three widely used family of test functions were proposed by Genz (1984). These test families were designed for

the problem of numerical integration. It is defined on $[0,1]^d$. Each of these families is given a name or attribute as follows:

(1) PRODUCT PEAK: $g_1(x) = \prod_{i=1}^{d}(c_i^{-2} + (x_i - w_i)^2)^{-1}$;

(2) CORNER PEAK: $g_2(x) = \left(1 + \sum_{i=1}^{d} c_i x_i\right)^{-(d+1)}$;

(3) CONTINUOUS: $g_3(x) = \exp\left(-\sum_{i=1}^{d} c_i \left|x_i - w_i\right|\right)$.

Different test functions can be obtained by varying the parameters $c = (c_1, c_2, \cdots, c_d)$ and $w = (w_1, w_2, \cdots, w_d)$. The parameter w acts as a shift parameter, and the difficulty of the functions is monotonically increasing with the $c_i > 0$. Our examples are for the dimension $d = 10$, and we use parameters c_i such that

$$\sum_{i=1}^{10} c_i = b_j, \tag{5.2.5}$$

where b_j is arbitrarily chosen for each family g_j, and w and c were generated independently and uniformly distributed in $[0,1]^d$. Then c was renormalized to satisfy (5.2.5).

In this paper, we test $g(x)$ for each test family with dimension $d = 10$. The value of b_j is given in Table 5.4.

Table 5.4: Value of b_j for Genz test functions

j	1	2	3
b_j	50	100	30

We apply the method of SOIR to evaluating the test functions PRODUCT PEAK, CORNER PEAK and CONTINUOUS. With the initial parameter $n_0 = 1 \times 10^6$ and $l = 30$, the numerical results of above three test functions are presented in Table 5.5. We also present the marginal histograms over importance sampling region over each of 10 dimensions for all of above three test functions, both on the initial stage and in the final stage (see in Figure 5.9—Figure 5.14).

Table 5.5: Numerical Results for Three Genz Test Functions

Numerical Result	PRODUCT PEAK	CORNER PEAK	Continuous
$\hat{I}$	2.1107×10^8	7.7750×10^{-12}	4.2725×10^{-7}
$\mathrm{Var}(\hat{I})$	4.1531×10^7	1.8574×10^{-12}	5.1425×10^{-8}

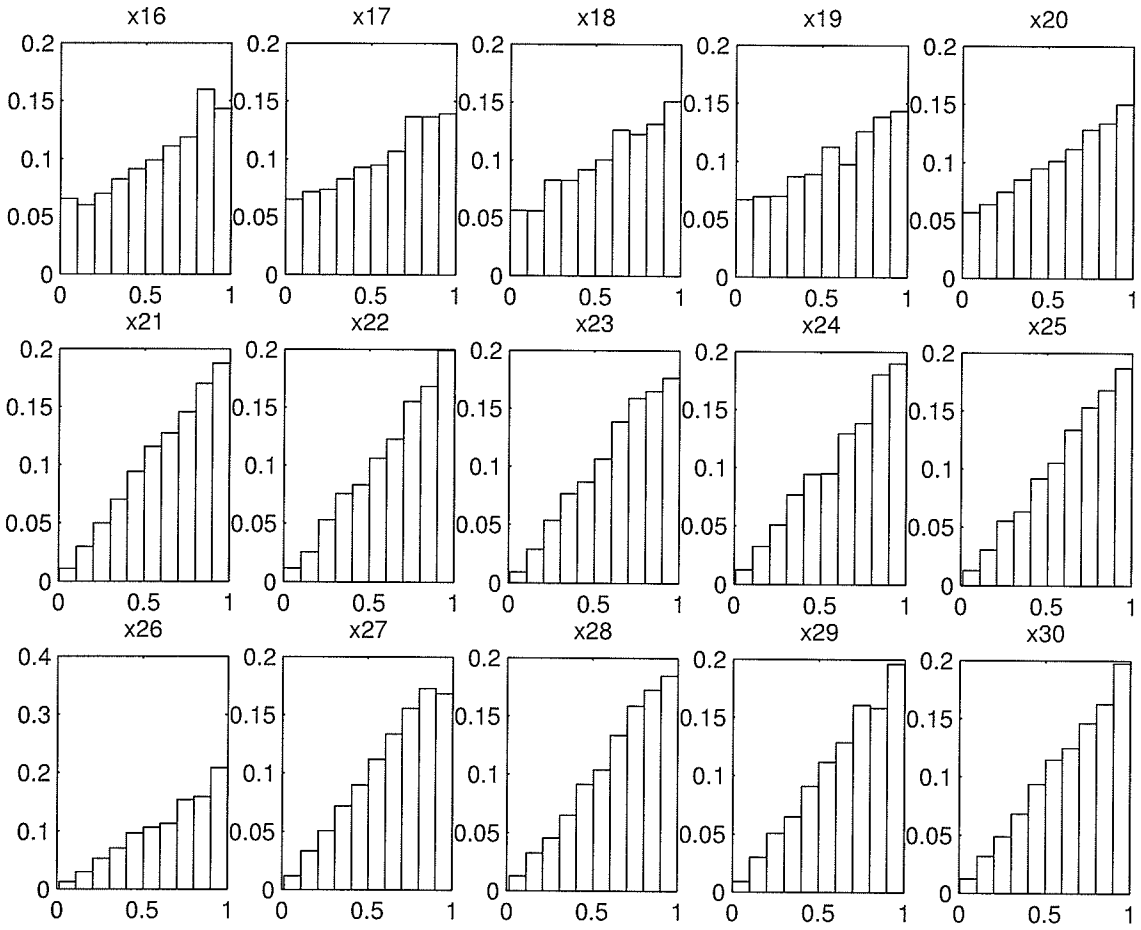

Figure 5.5: Histogram of marginal distribution of example 8 over importance region - part 2

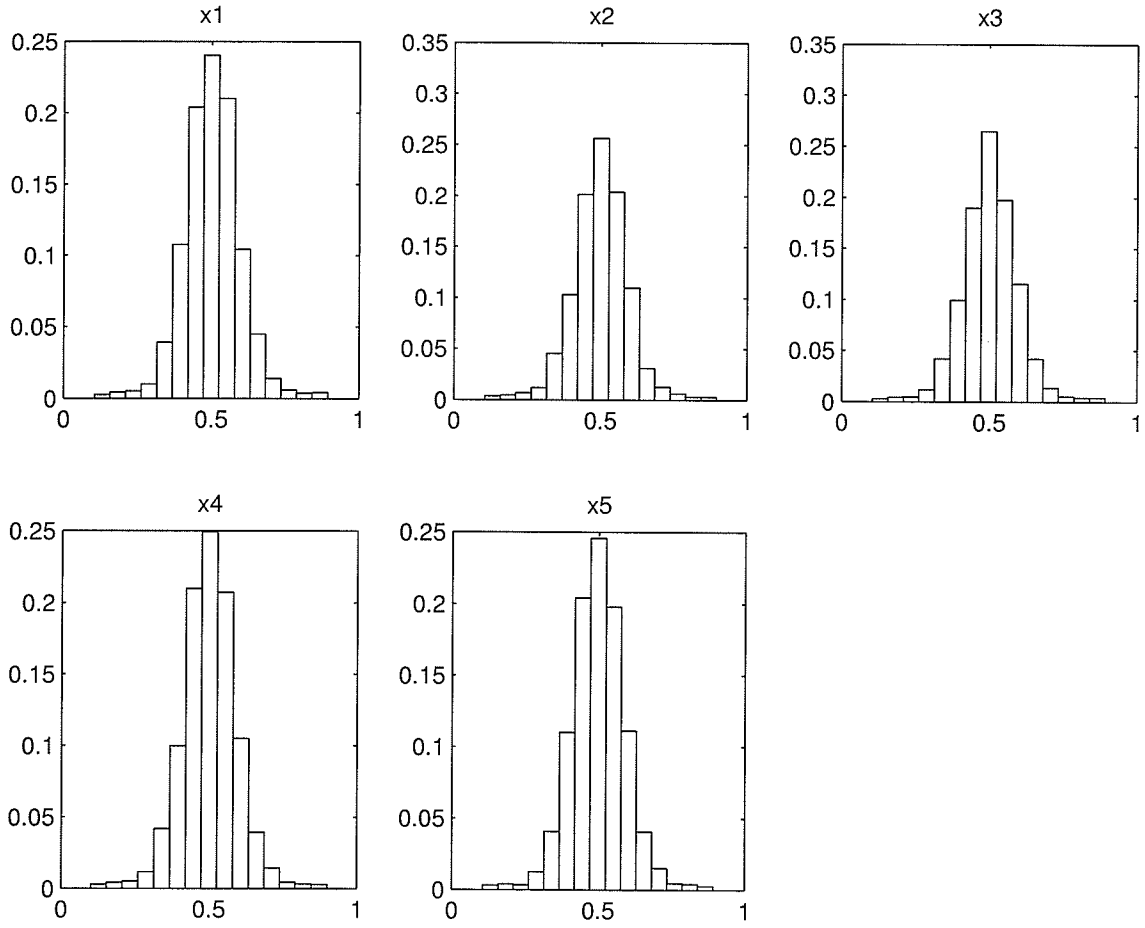

Figure 5.6: Histogram of marginal distribution of the five-dimensional mixture beta distribution after 4th iteration by using SOIR

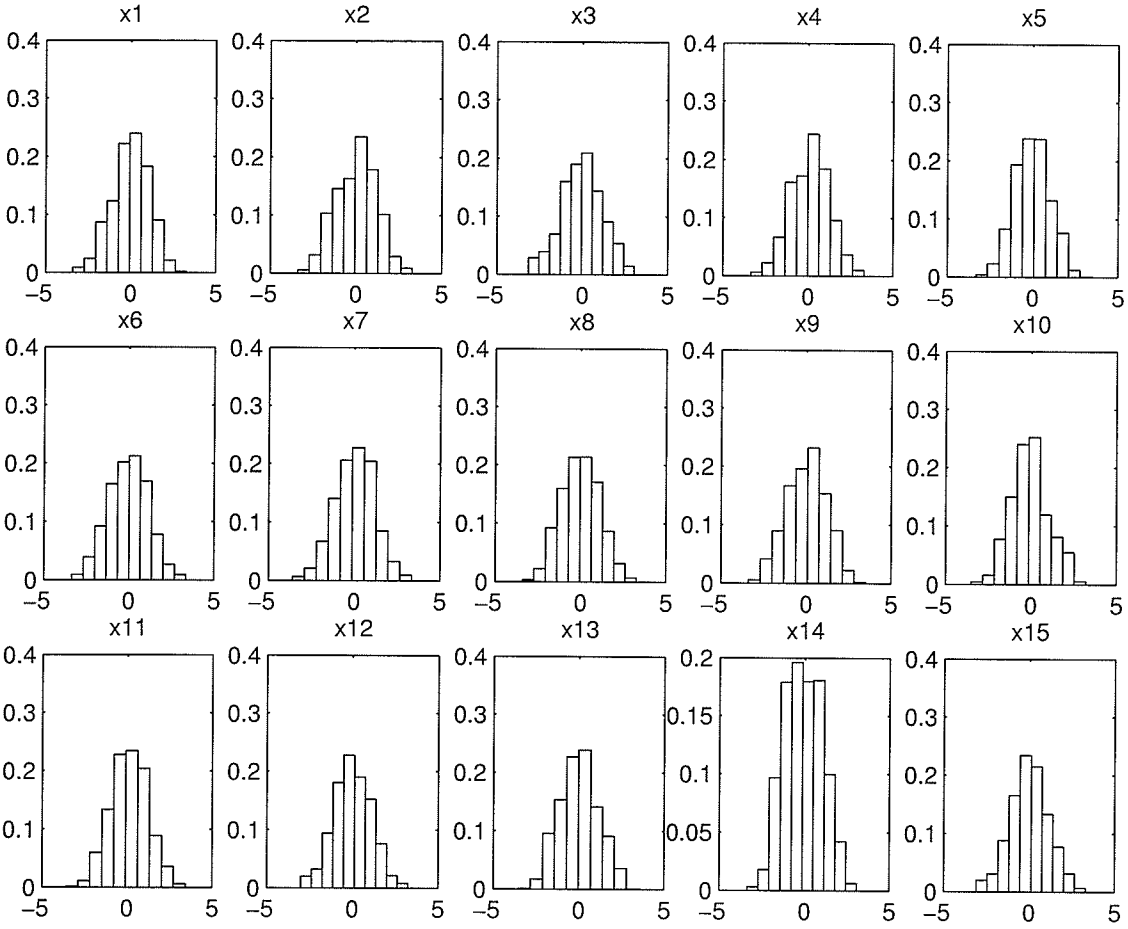

Figure 5.7: Histogram of marginal distribution of Capstick and Keister (1996) example over importance region - part 1

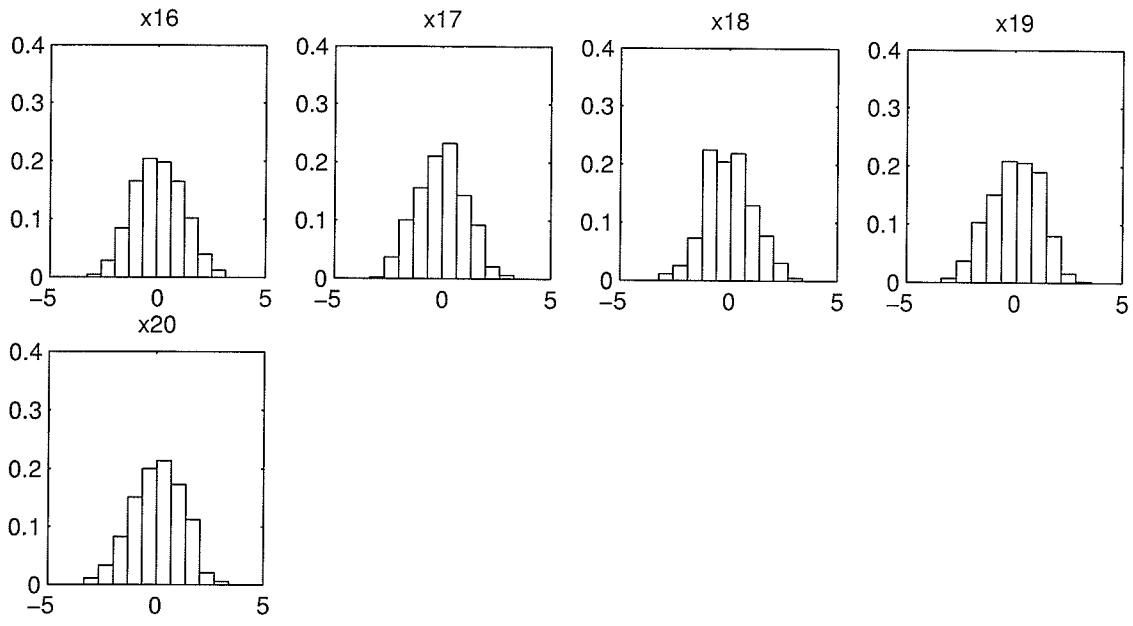

Figure 5.8: Histogram of marginal distribution of Capstick and Keister (1996) example over importance region - part 2

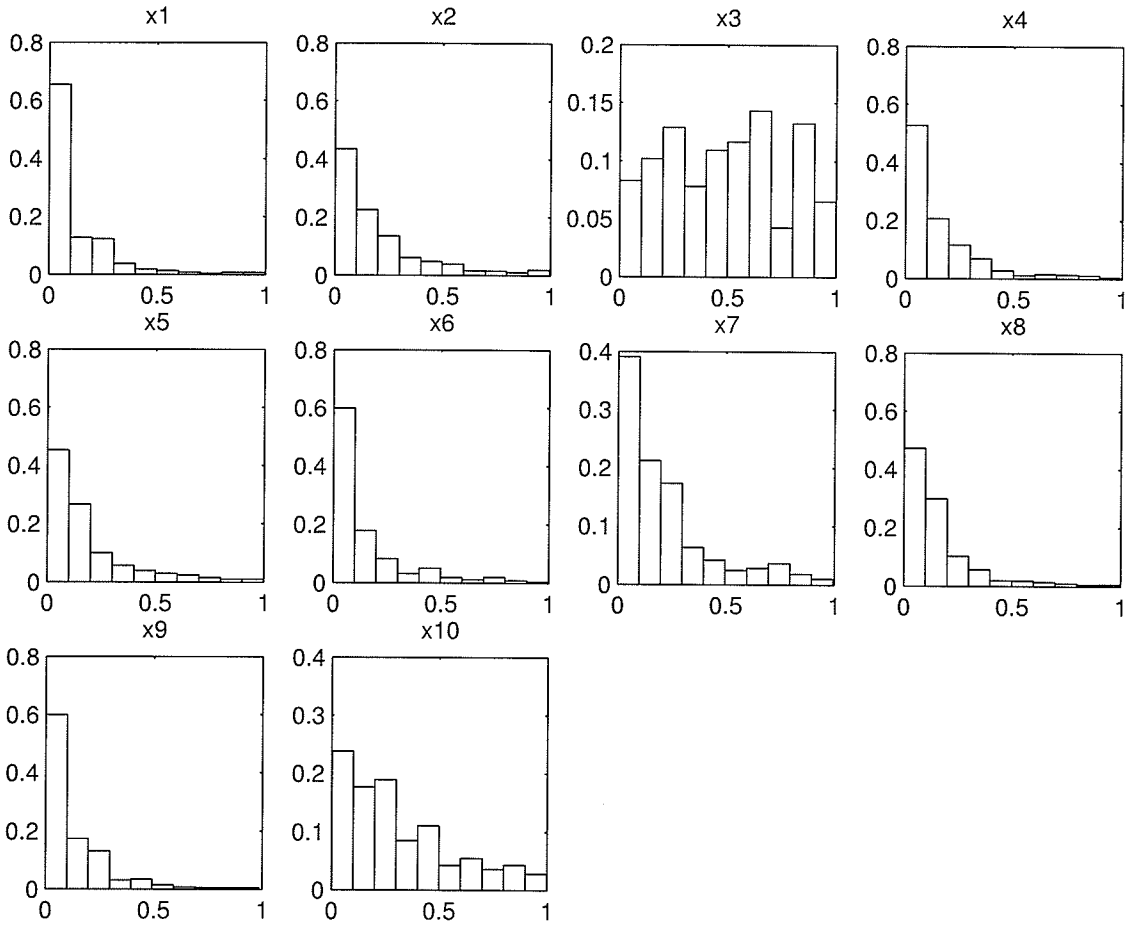

Figure 5.9: Histogram of marginal distribution of Genz Product Peak test function over initial importance region

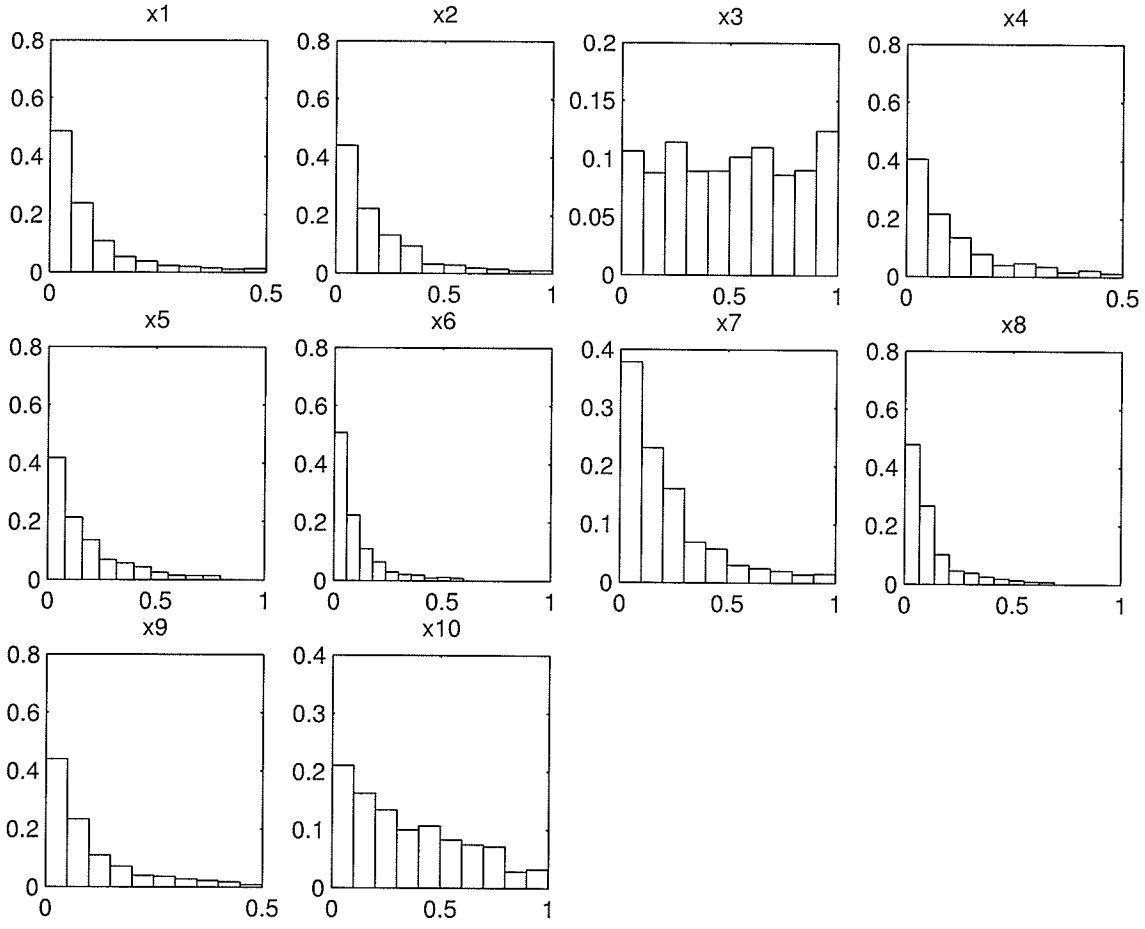

Figure 5.10: Histogram of marginal distribution of Genz Product Peak test function over final importance region

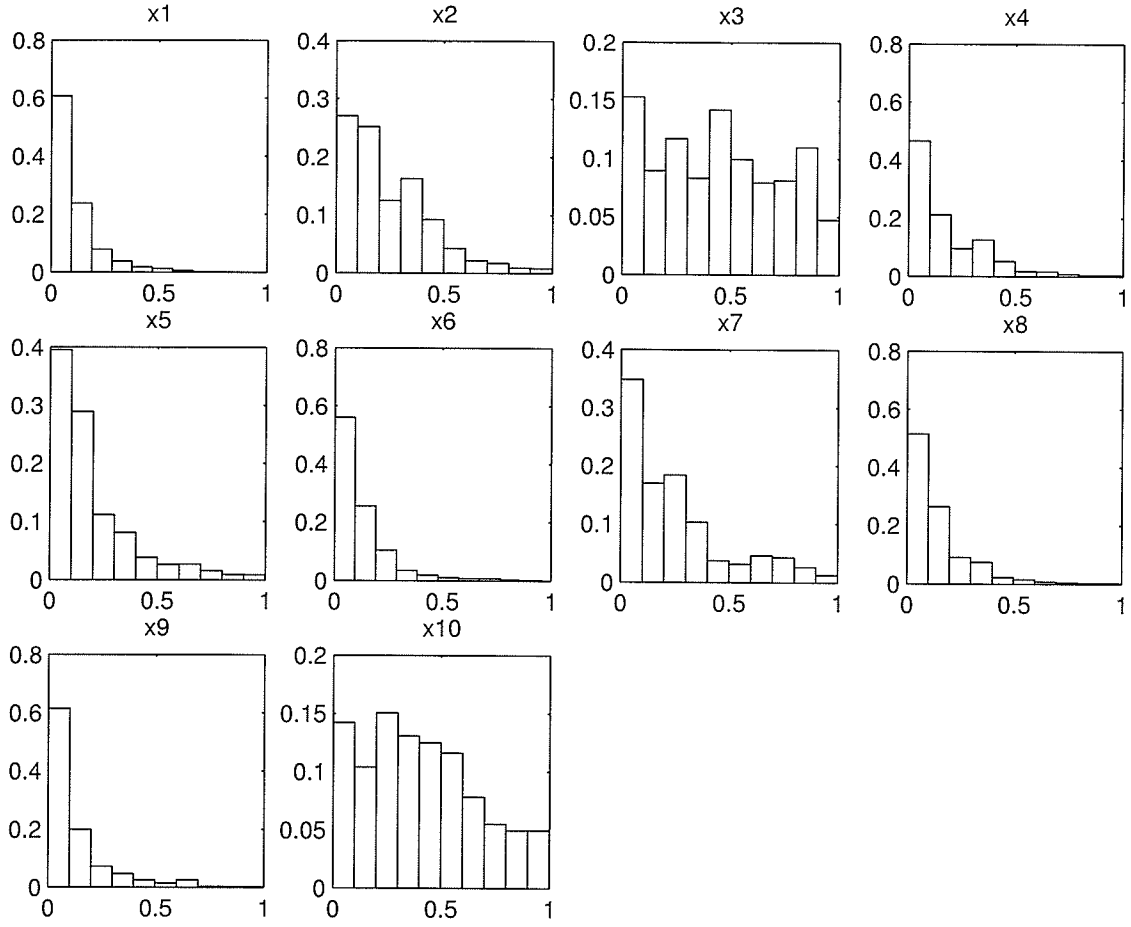

Figure 5.11: Histogram of marginal distribution of Genz Corner Peak test function over initial importance region

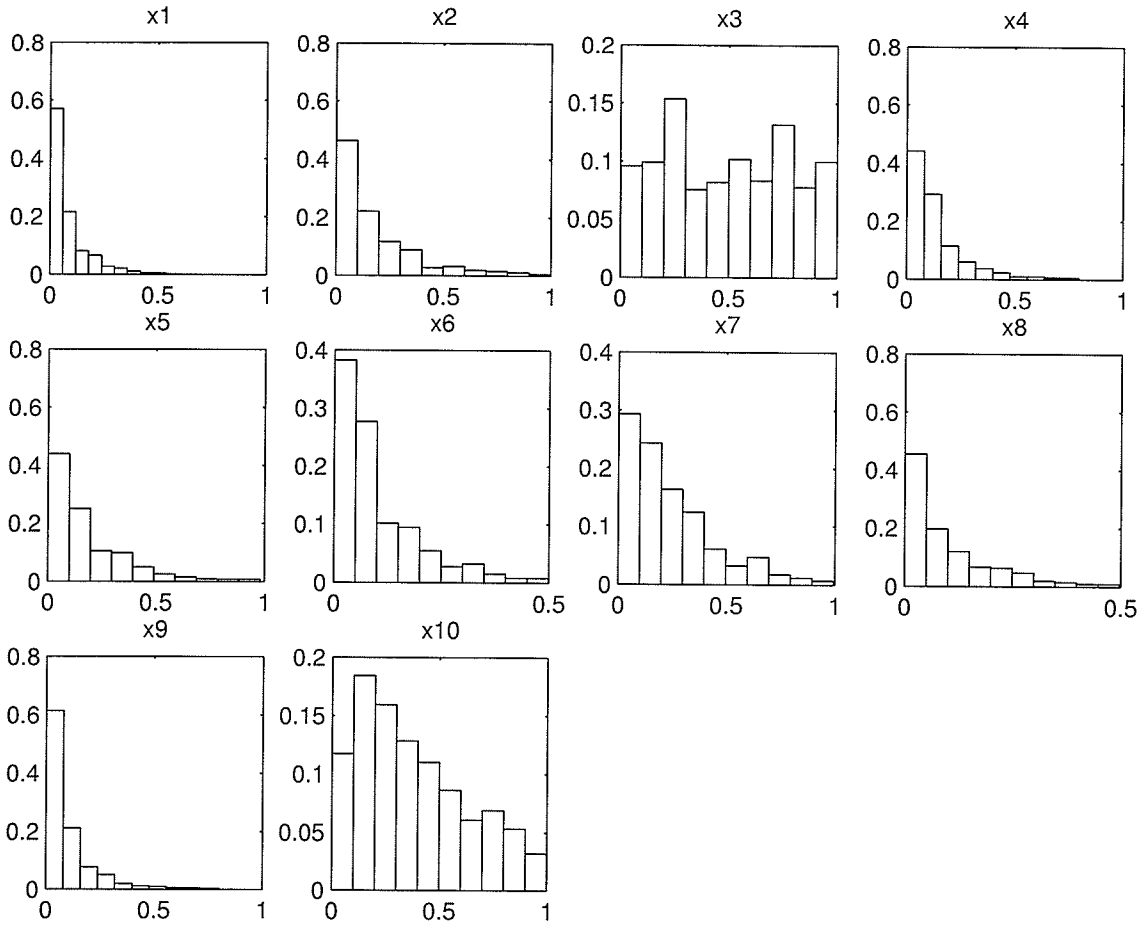

Figure 5.12: Histogram of marginal distribution of Genz Corner Peak test function over final importance region

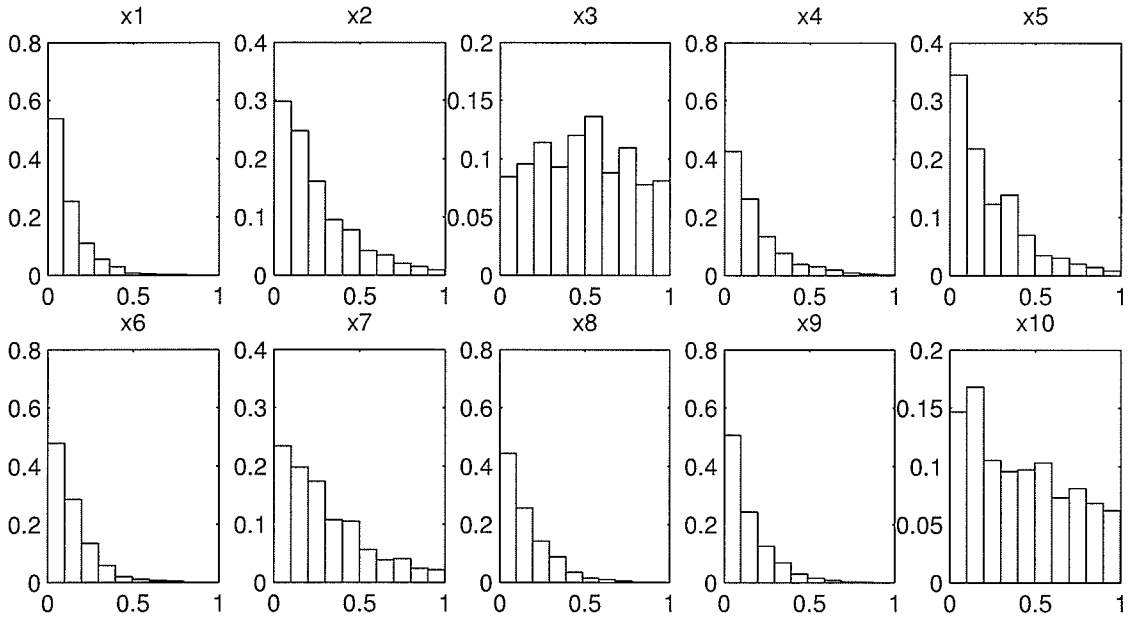

Figure 5.13: Histogram of marginal distribution of Genz Continuous test function over initial importance region

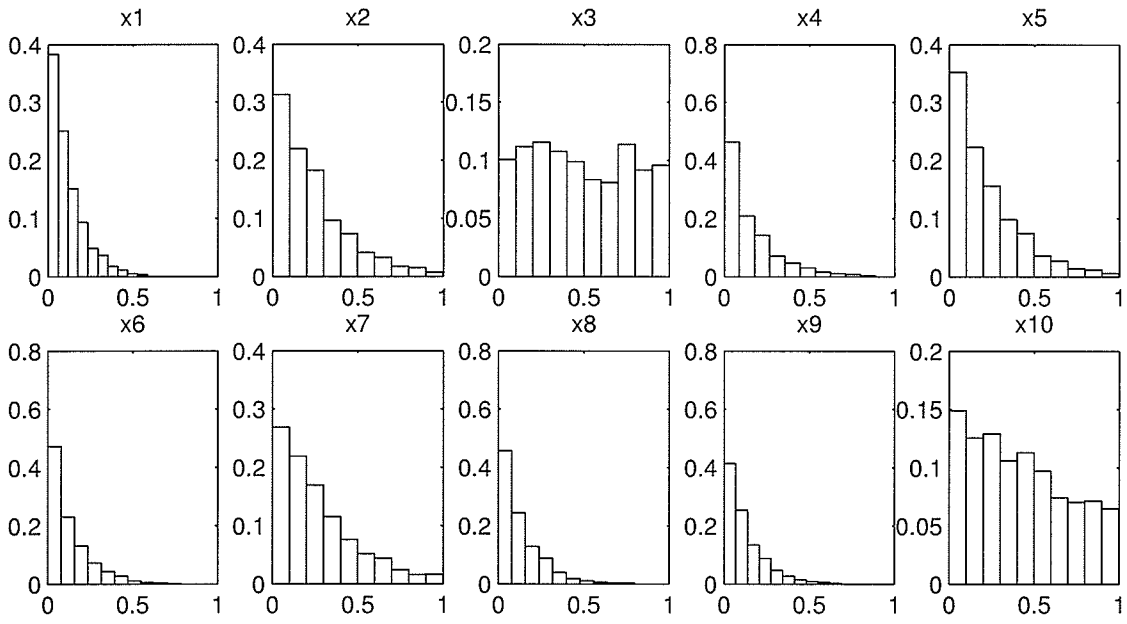

Figure 5.14: Histogram of marginal distribution of Genz Continuous test function over final importance region

Chapter 6

Summary and Further Research

If we survey in the development of numerical integration method retrospectively, we discern a rapid advancement in one-dimensional cases during the past.

For a number of reasons, numerical integration method in high dimensions have developed more slowly. Before the development of electronic digital computer, no practical use could be found for an integration method in higher dimensions. The availability of powerful computing devices and statistical softwares have however made possible the study of functions of multivariate distributions. New attempts at obtaining numerical integration methods in high-dimension are being made.

Among many integration techniques, importance sampling method is a popular tool for high-dimensional integration. The accuracy of its result, however, depends on the chosen importance density — a good choice may increase efficiency, but an inappropriate choice can cause devastating loss in accuracy. Moreover, in practice there is no guaranteed way to choose a good importance density. This problem

becomes more severe with the increasing complexity of the integrand, the number of dimensions, and the precision desired.

The method of sampling over importance region (SOIR) is developed to compare with numerous integrals estimated by known importance sampling methods in today's literature. After the formation of importance region of given random variables, the method of SOIR is to produce an initial acceptable estimator of given integral within reasonable computational time and cost. Furthermore, it is dimension-free and easy to implement. It is also applicable to many kinds of real functions arising in real life, including non-differentiable, and discontinuous functions.

For a given function of $g(x)$, the general construction of high-dimensional importance sampling region rule is a difficult task. Even though a large number of theoretical results are available today, the general building rule with safe coverage importance region still remains an art. The method of SOIR is mainly suitable for the functions with concave shapes. The value of $g(x)$ with respect to random variable X are highly concentrated at a hyper-cube subregion of the support of $g(x)$. However, for any convex shape functions, the method of SOIR is somewhat unsatisfactory because of the difficulty of developing an algorithm to build the appropriate importance region. Further research is required to solve such a problem in order to generalize the existing numerical integration algorithm.

Bibliography

[1] Au, S. K. and Beck, J. L. (2003) Important Sampling in High Dimensions. *Structural Safety.* **25**, 139-163.

[2] Bain, L. and Engelhardt, M. (1992) *Introduction to Probability and Mathematical Statistics* Second Edition, Duxbury Press. Belmont, California.

[3] Barthelmann, V, Novak., E., and Ritter, K. (2000) High Dimemsional Polynomial Interpolation on Sparse Grids. *Advances in Computational Mathematics.* **12**, 273-288.

[4] OH, M. and Berger, J. (1991) Adaptive importance Sampling in Monte Carlo Integration. *Journal. Statist. Computation and Simulation.* **41**, 143-168.

[5] OH, M. and Berger, J. (1993) Integration of Multimodal Functions by Monte Carlo Importance Sampling. *Journal of the American Statistical Association.* **88**, No. 422.

[6] Cheng, Jian. (2000) AISD-BN: An Adaptive Importance Sampling Algorithm for Evidential Reasoning in Large Bayesian Networks. *Journal of Artificial Intelligence Research.* **13**, 155-188.

[7] Evans, M. and Swartz, T. (1995) Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian integration Problems. *Statistical Science.* **10**, No.3, 254-272.

[8] Evans, M. and Swartz, T. (2000) *Approximation Integrals via Monte Carlo amd Deterministic Methods* First edition. Oxford University Press, New York.

[9] Fu, James C. and Wang, Liqun. (2002) A Random-Discretization Based Monte Carlo Sampling Method and Its Application. *Methodology and Computing in Applied Probability.* **4**, 5-25.

[10] Ford, B. and Thomasset, F. (1984) Tools, Methods and Languages for Scienctific and Engineering Computation. *Elsevier Science Publishers, England.*

[11] Givens, G. H. and Raftery, A. E. (1996) Local Adaptive Importance Sampling for Multivariate Densities With Strong Nonlinear Relationship. *Journal of the American Statistical Association.* **91**, No. 433.

[12] Glasserman, P., Heidelberger, P., and Shaharuddin, P. (1999) Asymptotically Optimal Importance Sampling and Stratification for Pricing path-Dependent Options. *Mathematical Finance.* **9**, No. 2, 117-152.

[13] Karaivanova, A. and Dimov, I. (1998) Error Analysis of an Adaptive Monte Carlo Method for Numerical Integration. *Mathematics and Computers in Simulation.* **47**, 201-213.

[14] Madras, N. and Piccioni, M. (1999) Importance Sampling for Families of Distributions. *The Annals of Applied Probability.* **9**, No. 4, 1202-1225.

[15] Novak, E., Ritter, K., Schmitt, R., and Steinbauer A. (1999) On an Interpolatory Method for High Dimensional Integration. *Journal of Computational and Applied Mathematics.* **112**, 215-228.

[16] Owen, A. and Zhou, Y. (2000) Safe and Effective Importance Sampling. *Journal of the American Statistical Association.* **95**, No. 449.

[17] Owen, A. (2003) The Dimension Distribution and Quadrature Test Functions. *Statistica Sinica.* **13**, 1-17.

[18] Schurer, R. (2003) A Comparison Between (Quasi-) Monte Carlo and Cubature Rule Based Methods for Solving High-Dimensional Integration problems. *Mathematics and Computers in Simulation.* **62**, 509-517.

[19] Snyder, W. C. (2000) Accuracy Estimation for Quasi-Monte Carlo simulations. *Mathematics and Computers in Simulation.* **54**, 131-143.

[20] Usabel, M. A. (1998) Application to Risk Theory of a Monte Carlo Multiple Integration Method. *Insurance Mathematics and Economics.* **23**, 71-83.

[21] Zhang P. (1996) Nonparametric Importance Sampling. *Journal of the American Statistical Association.* **91**, No. 435.