

Chapter 4

Monte Carlo integration

4.1 Introduction

The method of simulating stochastic variables in order to approximate entities such as

$$I(f) = \int f(x)dx$$

is called Monte Carlo integration or the Monte Carlo method. In applied engineering complicated integrals frequently surfaces and close form solutions are a rarity. In order to deal with the problem numerical methods and approximations are employed. By simulation and on the foundations of the *law of large numbers* is it possible to find good estimates for $I(f)$. Note here that "good" denotes close to the exact value in some specific sense.

4.1.1 Monte Carlo in probability theory

We will see how the Monte Carlo method for ordinary integrals extends to probability theory.

First recall the probabilistic concept of a expected value. If g is a function and X a stochastic variable with density function f_X then

$$\mathbb{E}[g(X)] = \int_{\Omega} g(x)f_X(x)dx.$$

Further, we see that calculating the expected value of $g(X)$ is actually equivalent to computing $I(f)$ for a suitable choice of g .

The foundations of Monte Carlo integration rests on the law of large numbers. Note first that a sequence of random variables, $\{X_n\}_{n=1}^{\infty}$, converge in the meaning of L^2 if $\mathbb{E}[X_n] \rightarrow \mu$ and $\text{Var}(X_n) \rightarrow 0$, for some $\mu \in \mathbb{R}$.

Here we give the L^2 law of large numbers for i.i.d.³ sequences. For a proof see e.g. [1] page 36.

Theorem 4.1.1. *Let X_1, X_2, \dots be i.i.d. with $\mathbb{E}[X_i] = \mu$, $\text{VAR}(X_i) = \sigma^2 \in (0, \infty)$. If $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ then $\bar{X}_n \rightarrow \mu$ in L^2 .*

The theorem says, given some constraints of the first two moments of $\{X_i\}_{i=1}^n$, the sample mean converges to a fixed number. This is very useful if we would like to estimate an unknown $I(f)$. In order to see this, construct i.i.d. stochastic variables X_1, \dots, X_n and a function g such that

$$\mathbb{E}[g(X_i)] = I(f), \quad (4.1)$$

i.e. find a function g such that $g(x)f_X(x) = f(x)$ for all x . Then will the arithmetic mean of $\{g(X_i)\}_{i=1}^n$ converge to a number, and that number is $I(f)$! So the complicated integral is almost a sum of random variables.

Further, If X_i have distribution function f_X then this condition corresponds to finding a g such that $g(x) = f(x)/f_X(x)$, since then

$$\mathbb{E}[g(X_i)] = \int g(x)f_X(x)dx = \int f(x)dx.$$

For proper integrals, i.e. integrals over a bounded interval, the most strait forward example is the method employed with variates of uniform distribution.

Definition 4.1. *A stochastic variable, X , is uniform(a, b)-distributed if its density function $f_X(x) = \frac{1}{b-a}$ for all $x \in [a, b]$ and zero elsewhere.*

Example 4.1. *For two arbitrary end points $a < b$ on the real line, consider a function $f : [a, b] \rightarrow \mathbb{R}$ such that $I(|f|)$ and $I(f^2)$ both are bounded. The boundedness conditions are there to guarantee that $\mathbb{E}[f(X_i)]$ and $\text{Var}(f(X_i))$ exist. Further, let $g(x) = (b-a)f(x)$ and X_1, \dots, X_n be i.i.d. uniform(a, b) distributed stochastic variables. The boundedness of f combined with $a, b \in \mathbb{R}$ gives us that both $I(|g|)$ and $I(g^2)$ are bounded so the boundedness conditions of the law of large numbers are fulfilled. Further let S_n denote the sample mean of $\{g(X_i)\}_{i=1}^n$ i.e. $S_n = \frac{1}{n} \sum_{i=1}^n g(x_i)$ where $\{x_i\}_{i=1}^n$ are observations of $\{X_i\}_{i=1}^n$ and note that*

$$E[g(X_i)] = \int_a^b (b-a)f(x)f_X(x)dx = I(f),$$

³i.i.d. means independent and identically distributed.

since $f_X(x) = \frac{1}{b-a}$ for $x \in [a, b]$. Then by the law of large numbers converges S_n (in L^2) to the entity of interest, $I(f)$. That means for n large, $S_n \approx I(f)$ and the complicated integral is approximated by a sum of observations of random variables which are simple to generate.

For improper integrals is the uniform distribution inadequate. But any distribution defined on the same set as the integral, with a corresponding g fulfilling condition (4.1), may be utilized. Further, Monte Carlo integration with i.i.d. uniform(0,1) distributed stochastic variables will here be denoted, *ordinary* Monte Carlo integration. Note that for this particular case then finding a function g that fulfills condition (4.1) is easy since $g(x) = f(x)$ does the job.

Further if $I(f)$ is an integral of higher dimension, i.e. $f : \mathbb{R}^k \rightarrow \mathbb{R}$, for some $k \geq 2$, then the same technique employed with random vectors instead of random variables will give the desired result.

4.2 Convergence and the central limit theorem

The Monte Carlo approximation converges by the law of large numbers, as $n \rightarrow \infty$, to the real value $I(f)$ of the integral. The convergence is in L^2 , probability or almost sure. Each of convergence meaning that there is never a guarantee that the approximation is so and so close $I(f)$, but that it becomes increasingly unlikely that it is not, as n increases. Mathematically this is formulated as the distribution of the estimand becomes more and more concentrated near the true value of $I(f)$. The performance of an estimator is measured by the spread of its distribution. To study the spread, or error, we use the *central limit theorem* (CLT), for a proof see e.g. [1] page 112,

Theorem 4.2.1. *Let X_1, X_2, \dots be i.i.d. with $\mathbb{E}[X_i] = \mu$, $\text{VAR}(X_i) = \sigma^2 \in (0, \infty)$. If $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ then*

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

converges in distribution to a Gaussian stochastic variable with zero mean and unity variance.

So by CLT, the sample mean of i.i.d. random variables with expected value μ and variance σ^2 is approximately $N(\mu, \sigma^2/n)$ -distributed. Note the remarkably property that eventhough the difference $(\bar{X}_n - \mu)/\sigma$ is enlarged by a big number, \sqrt{n} , the product is contained in some sense.

If $\{x_i\}_{i=1}^n$ are observations of i.i.d. uniform(0,1) random variables, then the ordinary Monte Carlo approximation $S_n = \frac{1}{n} \sum_{i=1}^n f(x_i)$ of the integral $I(f)$ satisfies

$$\mathbf{P} \left(a \frac{\sigma}{\sqrt{n}} < S_n - I(f) < b \frac{\sigma}{\sqrt{n}} \right) \approx \Phi(b) - \Phi(a),$$

where $\sigma^2 = \int (f(x) - I(f))^2 dx$. Here, making use of the Monte Carlo method again,

$$\sigma^2 = \int (f(x) - I(f))^2 dx \approx \frac{1}{n} \sum_{i=1}^n (f(x_i) - S_n)^2 = \frac{1}{n} \sum_{i=1}^n f(x_i)^2 - S_n^2 = \hat{\sigma}^2.$$

Note also, the above analysis shows that the error of the Monte Carlo method is of the order $1/\sqrt{n}$, regardless of the dimension of the integral.

Example 4.2. Let $f(x) = \frac{4}{1+x^2}$, and employ ordinary Monte Carlo integration to compute the integral

$$I(f) = \int_0^1 f(x) dx.$$

The integral is then approximated with

$$S_n = \frac{1}{n} \sum_{i=1}^n \frac{4}{1+x_i^2},$$

where x_1, \dots, x_n are observations from i.i.d. uniform(0,1) distributed random numbers. A computer program for this could look as follows:

```
Est=0, Est2=0
For 1 to n
    Generate a uniform distributed random variable x_i.
    Compute y=4/(1+x_i^2)
    Est=Est+y and Est2=y^2+Est2
End
Est=Est/n and Est2=Est2/n
std=sqrt(Est2-Est^2)
```

4.3 Variance reduction

Since variance of S_n in some sense denotes the estimators performance, is it a central issue to find estimators with small variance. One way to reduce variance of an estimand is to employ *variance reduction* techniques. Where the idea basically is to transform the original observations by some transformation that conserves expected value but reduces the variance of the estimand.

It should be noted that a badly performed attempt to variance reduction, at worst leads to a larger variance, but usually nothing worse. Therefore, there is not too much to lose on using such techniques. And it is enough to feel reasonably confident that the technique employed really reduces the variance: There is no need for a formal proof of that belief!

There are a couple of standard techniques of variance reduction. The techniques often carry fancy names, but the ideas behind are relative strait forward.

4.3.1 Importance sampling

By chosing a distribution function of the random variables such that the density of the sampling points are close to the shape of the integrand, the variance of the sample mean decreases. This method of variance reduction is called importance sampling.

First notice that

$$I(f) = \int f(x)dx = \int \frac{f(x)}{p(x)}p(x)dx, \quad (4.2)$$

so if we select p to be a probability density function, we may, as an alternative to ordinary Monte Carlo integration, compute an approximation of $I(f)$ by the sample mean of $f(X_i)/p(X_i)$, where $\{X_i\}_{i=1}^n$ are i.i.d. random variables with probability density function p . An Monte Carlo approximation of the integral $I(f)$ conducted with importance sampling is

$$S_n = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i)}{p(x_i)},$$

where x_1, \dots, x_n are obseravtions of i.i.d. random variables with probability density function $p(x)$.

The variance of $f(X_1)/p(X_1)$ is estimated as before with

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{f(x_i)}{p(x_i)} \right)^2 - S_n^2.$$

If the shape of p is close to f then will the ratio f/p be close to a constant and thus will the variance be small, which is the endeavored property.

Example 4.3. *Continued example. Instead of using ordinary Monte Carlo integration in previous example, let X_1, \dots, X_n be i.i.d. stochastic variables with density function $f_X(x) = \frac{1}{3}(4 - 2x)$ for $x \in [0, 1]$ and zero elsewhere. Note here that indeed f_X is a density function since it is non-negative and $\int_{-\infty}^{\infty} f(x)dx = 1$. Further, in some sense is the shape of f_X close to f , see Figure 4.1. And the Monte Carlo approximation of $I(f)$ is*

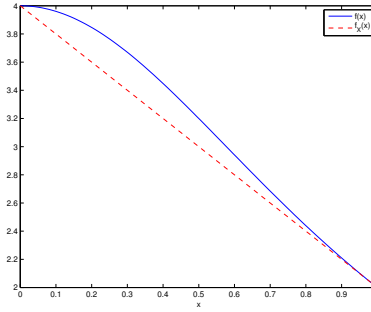


Figure 4.1: The functions $f(x)$ and (a scaled version of) $f_X(x)$.

$$S_n = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i)}{f_X(x_i)},$$

for observations x_1, \dots, x_n of the specific distribution. By the shape similarities of f and f_X is f/f_X not so fluctant over $x \in [0, 1]$ and thus is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{f(x_i)}{p(x_i)} \right)^2 - S_n^2$$

small.

Assume we have a stochastic variable, X , with density p such that $p(x) = cf(x)$ for all x and some $c \in \mathbb{R}$, and we know $\mathbb{E}[X]$. Then $f/p = c$ and $\text{Var}(S_n) = \frac{1}{n} \sum_{i=1}^n c^2 - S_n^2 = 0$. So $I(f) = I(p)/c$ which then is known.

4.3.2 Control variates

An alternative to importance sampling, is to employ a *control variate* p , which is a function that is close to the integrand, f , and with a known value

$I(p)$ of the integral. By linearity of integrals then

$$I(f) = \int f(x)dx = \int (f(x)-p(x))dx + \int p(x)dx = \int (f(x)-p(x))dx + I(p),$$

so if p is close to f then the first term on the righthand side is small. By Monte Carlo integrate only the indeterminated part, the variance is reduced. If S_n is the Monte Carlo approximation then

$$S_n = \frac{1}{n} \sum_{i=1}^n (f(x_i) - p(x_i)) + I(p),$$

and the variance of S_n originates only from the difference $f - p$.

Example 4.4. *Continued example. Instead of using importance sampling Monte Carlo integration in the exhausted example, note that if $p(x) = 4 - 2x$ for $x \in [0, 1]$ and zero elsewhere then $I(p) = 3$. Monte Carlo integration of $I(f)$ employed with this control variate is then*

$$S_n = I(p) + \frac{1}{n} \sum_{i=1}^n f(X_i) - p(X_i).$$

In analogy with importance sampling, since $p(x) \approx f(x)$ then is $f(x) - p(x)$ not so fluctant over $x \in [0, 1]$ and thus is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(f(x_i) - p(x_i) + I(p) \right)^2 - S_n^2$$

small.

4.3.3 Antithetic variates

Another technique to reduce variance, is to utilize *antithetic variates*, where simulated variates are recycled. In contrast to ordinary Monte Carlo integration, here the property of dependence is utilized. The motivation for this technique is, for stochastic variables X and Y then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

If covariance, of X and Y , is negative then is the variance of $X + Y$ less than the sum of each variates variance.

Example 4.5. Let $f : [0, 1] \rightarrow \mathbb{R}$ be a monotone function of one variable (i.e., f is either increasing or decreasing). In order to approximate the integral $I(f)$ using observed i.i.d. uniform(0,1) random numbers $\{x_i\}_{i=1}^n$, then a Monte Carlo integration with antithetic variables is

$$S_n = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i)}{2} + \frac{1}{n} \sum_{i=1}^n \frac{f(1-x_i)}{2},$$

where the variance of S_n is estimated by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{f(x_i)}{2} + \frac{f(1-x_i)}{2} \right)^2 - S_n^2,$$

which is less than the variance of the estimand of the ordinary Monte Carlo approximation.

In the above example, the random variable, $1 - X$ has the same distribution as X , but are negatively correlated. If X is big then $1 - X$ is bound to be small. Further if f monotone the non-positive correlation holds for $f(X)$ and $f(1 - X)$. In theory, any transformation, $T : [0, 1] \rightarrow [0, 1]$, is possible to employ to the ordinary Monte Carlo integration. Further if $f(X)$ and $f(T(X))$ are negatively correlated then is the variance of

$$\frac{1}{2n} \sum_{i=1}^n f(X_i) + f(T(X_i))$$

less than the corresponding entity for ordinary Monte Carlo integration.

Since, $1 - x_i$ is an equally good observation of a uniform(0,1) variable as x_i and further are x_i and $1 - x_i$ negatively correlated. This, in turn, entails that $f(x_i)$ and $f(1 - x_i)$ are negatively correlated, since f is monotone. Thus is the variance of S_n

4.3.4 Stratified sampling

Often the variation of the function f that is to be integrated varies over different parts of the domain of integration. In that case, it can be fruitful to use *stratified sampling*, where the domain of integration is divided into smaller parts, and use Monte Carlo integration on each of the parts, using different sample sizes for different parts.

Phrased mathematically, we partition the integration domain $M = [0, 1]^d$ into k regions M_1, \dots, M_k . For the region M_j we use a sample of size n_j of

observation $\{x_{ij}\}_{i=1}^{n_j}$ of a random variable X_j with a uniform distribution over M_j . The resulting Monte Carlo approximation S_n of the integral $I(f)$ becomes

$$S_n = \sum_{j=1}^k \frac{\text{vol}(M_j)}{n_j} \sum_{i=1}^{n_j} f(x_{ij})$$

In order for stratified sampling to perform optimal, one should try to select $n_j \sim \text{vol}(M_j)\sigma_{M_j}(f)$.

Example 4.6. Let $f : [-1, 1] \rightarrow \mathbb{R}$ be such that $f(x) = \sin(\pi/x)$ for $x \in [-1, 0]$, $f(x) = 1/2$ for $x \in [0, 1]$ and zero elsewhere and consider to compute $I(f)$. Note here that $f(x)$ is fluctuating heavily for negative x while constant for positive, see Figure 4.2. By splitting the integration interval in these two

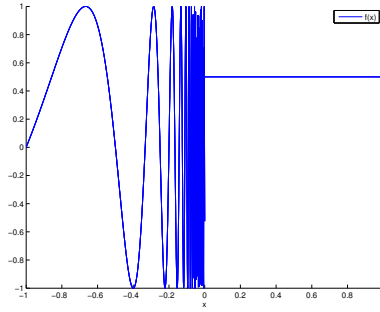


Figure 4.2: The function $f(x)$.

parts, employing ordinary Monte Carlo integration on them separately, the technique of stratified sampling is utilized.

4.4 Simulation of random variables

Since Monte Carlo integration is based on converging sums of stochastic variables we need an easy way of generating these random variables.

4.4.1 General theory for simulation of random variables

The following technical lemma is a key step to simulate random variables in a computer:

Lemma 4.1. *For a distribution function F , define the generalized right-invers F^{\leftarrow} by*

$$F^{\leftarrow}(y) \equiv \min\{x \in (0, 1) : F(x) \geq y\} \quad \text{for } y \in (0, 1).$$

We have

$$F^{\leftarrow}(y) \leq x \Leftrightarrow y \leq F(x).$$

Proof. ⁴For $F(x) < y$ there exists an $\epsilon > 0$ such that $F(x) < y$ for $z \in (-\infty, x + \epsilon]$, as F is non-decreasing and continuous from the right. This gives

$$F^{\leftarrow}(y) = \min\{z \in (0, 1) : F(z) \geq y\} > x.$$

On the other hand, for $x < F^{\leftarrow}(y)$ we have $F(x) < y$, since

$$F(x) \geq y \Rightarrow F^{\leftarrow}(y) = \min\{z \in (0, 1) : F(z) \geq y\} \leq x.$$

Since we have shown that $F(x) < y \Leftrightarrow x < F^{\leftarrow}(y)$, it follows that $F^{\leftarrow}(y) \leq x \Leftrightarrow y \leq F(x)$. \square

From a uniform(0,1) random variable a random variable with any other desired distribution can be simulated, at least in theory:

Theorem 4.1. *If F is a distribution function and ξ a uniform(0,1) random variable, then $F^{\leftarrow}(\xi)$ is a random variable with distribution function F .*

Proof. Since the uniformly distributed random variable ξ has distribution function $F_{\xi}(x) = x$ for $x \in [0, 1]$, Lemma 4.1 shows that

$$F_{F^{\leftarrow}(\xi)}(x) = \mathbf{P}\{F^{\leftarrow}(\xi) \leq x\} = \mathbf{P}\{\xi \leq F(x)\} = F_{\xi}(F(x)) = F(x). \quad \square$$

When using Theorem 4.1 in practice, it is not necessary to know an analytic expression for F^{\leftarrow} : It is enough to know how to calculate F^{\leftarrow} numerically.

If the distribution function F has a well-defined ordinary inverse F^{-1} , then that inverse coincides with the generalized right-inverse $F^{\leftarrow} = F^{-1}$.

Corollary 4.1. *Let F be a continuous distribution function. Assume that there exists numbers $-\infty \leq a < b \leq \infty$ such that*

- $0 < F(x) < 1$ for $x \in (a, b)$;
- $F : (a, b) \rightarrow (0, 1)$ is strictly increasing and onto.

⁴This proof is not important for the understanding of the rest of the material.

Then the function $F : (a, b) \rightarrow (0, 1)$ is invertible with inverse $F^{-1} : (0, 1) \rightarrow (a, b)$. Further, if ξ is a $\text{uniform}(0, 1)$ random variable, then the random variable $F^{-1}(\xi)$ has distribution function F .

Corollary 4.1 might appear to be complicated, at first sight, but in practice it is seldom more difficult to make use of it than is illustrated in the following example, where F is invertible on $(0, \infty)$ only:

Definition 4.2. A stochastic variable, X , is $\exp(\lambda)$ -distributed if its cumulative distribution function $F_X(x) = 1 - e^{-\lambda x}$ for $x > 0$ and zero elsewhere.

Example 4.7. The inverse of an $\exp(\lambda)$ -distribution is

$$F^{-1}(y) = -\lambda^{-1} \ln(1 - y) \quad \text{for } y \in (0, 1).$$

Hence, if ξ is a $\text{uniform}(0, 1)$ random variable, then Corollary 4.1 shows that

$$\eta = F^{-1}(\xi) = -\lambda^{-1} \ln(1 - \xi) \quad \text{is } \exp(\lambda)\text{-distributed.}$$

This gives us a recipe for simulating $\exp(\lambda)$ -distributed random variables in a computer.

4.4.2 Simulation of normal distributed random variables

It is possible to simulate Normal distributed random variables by an application of the above Theorem 4.1. But since there are no closed form expressions for the inverse normal distribution function only numerical solutions exist. The standard way of getting round this intractability is to simulate normal distributed stochastic variables by the Box-Müller algorithm.

Theorem 4.2 (Box-Müller). If ξ and η are independent $\text{uniform}(0, 1)$ random variables, then we have

$$Z \equiv \mu + \sigma \sqrt{-2 \ln(\xi)} \cos(2\pi\eta) \quad N(\mu, \sigma^2)\text{-distributed}$$

Proof. ⁵For N_1 and N_2 , independent $N(0, 1)$ -distributed, the two-dimensional vector (N_1, N_2) has radius $\sqrt{N_1^2 + N_2^2}$ that is distributed as the square-root of a $\chi(2)$ -distribution. Moreover, a $\chi(2)$ -distribution is the same thing as an $\exp(1/2)$ -distribution.

By symmetry, the vector (N_1, N_2) has argument $\arg(N_1, N_2)$ that is uniformly distributed over $[0, 2\pi]$.

⁵This proof is not important for the understanding of the rest of the material.

Adding things up, and using Example 4.7, it follows that, for ξ and η independent uniform(0,1) random variables,

$$(N_1, N_2) =_{\text{distribution}} \sqrt{-2 \ln(\xi)} (\cos(2\pi\eta), \sin(2\pi\eta)). \quad \square$$

4.4.3 Simulation of γ -distributed random variables

As for normal random variables, by an application of Theorem 4.1 it is possible but intracable to simulate γ distributed random variables by the inverse of the distribution function. For integer values of the shape parameter, k , is a strait forward method of generating this data by summing k independent $\exp(\lambda)$ distributed variables.

Theorem 4.3 (Erlang distribution). *If $\{\xi_i\}_{i=1}^k$ are i.i.d. $\exp(\lambda)$ distributed random variables, then we have*

$$\sum_{i=1}^k \xi_i \sim \gamma(k, \lambda) - \text{distributed}$$

Proof. ⁶Let $f_{\exp(\lambda)}$ denote the density of ξ_i , then is $(\mathcal{F}f_{\exp(\lambda)})^k$ the Fourier transform of the density of $\sum_{i=1}^k \xi_i$. Further, $\mathcal{F}(f_{\exp(\lambda)})(\omega) = \frac{\lambda}{\lambda + 2\pi i\omega}$ and by tedious algebra

$$\mathcal{F}^{-1}\left(\left(\frac{\lambda}{\lambda + 2\pi i\omega}\right)^k\right)(x) = f_{\gamma(k, \lambda)}(x),$$

where $f_{\gamma(k, \lambda)}(x)$ denotes the density of a γ distributed random variable with parameters (k, λ) . \square

4.5 Software

The computer assignment is to be done in C. Since plotting figures is troublesome for non native C-programmers then use matlab for graphical aid. It may be very helpful having an figure to connect to when employing the variance reduction. Further, note that C is very fast so do not be modest in terms of number of generated random variables.

Below you will find the embryo of a C program, where the Box-Müller algorithm is incorporated.

⁶This proof is not important for the understanding of the rest of the material.

```

/* C-program with functions for generation of
N(mu,sigma^2)-distributed random variables */
#include <math.h> //for mathfunctions
#include <stdio.h>
#include <stdlib.h>
#include <time.h> // for timing

//function for generate normal random variable (mu,sigma) (Box-Muller)
double normrnd(double mu,double sigma)
{
    double chi,eta;
    chi=drand48();
    eta=drand48();
    //M_PI is a constant in math.h
    return mu+sigma*sqrt(-2*log(chi))*cos(2*M_PI*eta);;
}
main()
{
    // Define parameters and conduct Monte Carlo simulations

    return;
}

/* To compile the program in a terminal execute:
gcc -o app filename.c -lm
To run the program in a terminal execute: ./app
*/

```

4.6 Computer assignment

When doing variance reduction, please explain why a particular method actually reduces variance, when it is more/less effective and why it has been more/less effective in your particular situation. In order to pass the lab you will need 3 points.

Assignment 1 (4p): Let $f(x) = \frac{4}{1+x^2}$ for $x \in [0, 1]$ and zero elsewhere.

- Use ordinary Monte Carlo integration to approximate the integral $I(f)$ numerically. Do this for several "sample sizes" n , for example $n = 10^5, 10^6, 10^7, \dots$

By the ingenious variable substitution $x = \tan(\theta)$, observing that $dx = (\tan^2(\theta) + 1)d\theta$ by the trigonometric one, it entails that $I(f) = \pi$.

- Assume the exact value of π is unknown and construct 95% confidence intervals of π by the Monte Carlo estimates. Comment on the relation between the confidence intervals and the exact entity.
- Pick three variance reduction techniques (whichever you want) and re-calculate the integral by applying those. Do this for the same n values as before. Do you get more accurate estimates of π ? Why or why not?

Do the above by writing a function in C that takes in one "sample size" n and returns the estimate of the integral and the variance of the estimate. You can write a separate function for the variance reduction or change the original one so that it returns both the simple estimate and the ones obtained through variance reduction.

Assignment 2 (2p): The Monte Carlo method is not applicable for all set ups. Let

$$f(x, y) = \left| \frac{\sin(xy)}{x - \frac{1}{2}} \right|,$$

for all $(x, y) \in [0, 1]^2$ and zero elsewhere.

- Compute estimations of $I(f) = \int_0^1 \int_0^1 f(x, y) dx dy$ by the Monte Carlo method. Are the estimates accurate? Why or why not?

Assignment 3 (4p): Monte Carlo integration for improper integrals. Let

$$f(x) = \exp(-|x|) \sin(x) x^3,$$

for $x \in \mathbb{R}$.

- Compute estimations of $I(f) = \int_{-\infty}^{\infty} f(x) dx$ by Monte Carlo integration.

Comment: The probability density function of an exponential distributed stochastic variable is, $f(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$ and zero elsewhere. The probability density function of an gamma distributed stochastic variable is, $f(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp(-x/\theta)$ for $x \geq 0$ and zero elsewhere.

- Pick a variance reduction technique and re-calculate the integral. Do this for the same n values as before. Do you get more accurate estimates? Why or why not?

Bibliography

- [1] DURRETT, R. (2010). *Probability: Theory and Examples*, 4th edn. Cambridge University Press UPH, p 60, Shaftesbury Road, Cambridge, CB2 8BS, UK.