Chapter 5

Monte Carlo integration

5.1 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. This is desirable in applied mathematics, where complicated integrals frequently arises and closed 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* it is possible to find good estimates for $I(f)$. Note here that "good" means "close to the exact value" in some specific sense.

There are other methods to approximate an integral such as $I(f)$. One common way is the *Riemann sum*, where the integral is replaced by the sum of small n intervals in one dimension. In order to calculate a d-dimensional integral, it is natural to try to extend the one-dimensional approach. When doing so, the number of times the function f has to be calculated increases to $N = (n+1)^d \approx n^d$ times, and the approximation error will be proportional to $n^{-2} \approx N^{-2/d}$. One key advantage of the Monte Carlo method to calculate integrals numerically, is that it has an error that is proportional to $n^{-1/2}$, regardless of the dimension of the integral.

5.2 Monte Carlo Integration

Consider the d-dimensional integral

$$I = \int f(x)dx = \int_{x_1=0}^{x_1=1} \cdots \int_{x_d=0}^{x_d=1} f(x_1, \ldots, x_d)dx_1 \ldots dx_d$$

of a function f over the unit hypercube $[0,1]^d = [0,1] \times \ldots \times [0,1]$ in $\mathbb{R}^d$. Notice that the integral can be interpreted as the expectation $\mathbf{E}\{f(X)\}$ of the random variable $f(X)$, where X is an $\mathbb{R}^d$-valued random variable with a uniform distribution over $[0,1]^d$, meaning that the components $X_1, \ldots, X_d$ are independent and identically uniformly distributed over $[0,1]$, i.e., $X_1, \ldots, X_d$ are random numbers.

The *Monte Carlo approximation* of the integral is given by

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

where $\{x_i\}_{i=1}^n$ are independent observations of X, i.e. independent random observations of a $\mathbb{R}^d$-valued random variable, the components of which are random numbers. The extension to a hyperrectangle instead of $[0,1]^d$ is straightforward.

5.2.1 Monte Carlo in probability theory

We will see how to use the Monte Carlo method to calculate integrals. Since probabilities and expectations can in fact be described as integrals, it is quite immediate how the Monte Carlo method for ordinary integrals extends into probability theory.

Lets start by recalling the probabilistic concept of an expected value. If g is a function and X a stochastic variable with density function f_X then

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

where Ω is the support of f_X. Note that calculating the expected value of $g(X)$ is actually equivalent to computing $I(f)$ for a suitable choice of g.

5.2.2 Convergence and the Law of large numbers

The foundations of Monte Carlo integration rests on the law of large numbers. In fact, the above approximation converges, by the law of large numbers, as $n \to \infty$, to the real value I of the integral. The convergence is in the probabilistic sense, that there is never a guarantee that the approximation is so and so close I, but that it becomes increasingly unlikely that it is not, as $n \to \infty$.

Note first that a sequence of independent random variables, $\{X_n\}_{n=1}^\infty$, converge in the meaning of L^2 if $\mathbb{E}[X_n] \to \mu$ and $\mathrm{Var}(X_n) \to 0$, for some $\mu \in \mathbb{R}$. We have then, the law of large number in L^2 for independent and identically distributed (i.i.d.) sequences:

Theorem 5.2.1 (Law of Large Numbers)**.** *Let $X_1, X_2, ...$ be i.i.d. with $\mathbb{E}[X_i] = \mu \in \mathbb{R}$, $Var(X_i) = \sigma^2 \in (0, \infty)$. If $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ then $\bar{X}_n \to \mu$ in L^2.*

The theorem states that given the constraints of the first two moments of X_i (i.e.: $\mathbb{E}[X_i] = \mu \in \mathbb{R}$ and $\mathrm{Var}(X_i) = \sigma^2 \in (0, \infty)$) then the sample mean converges to a fixed number. This fact allows us to find $I(f)$. In order to see this, construct a sequence of i.i.d. stochastic variables $X_1, ..., X_n$ and a function g such that

$$\mathbb{E}[g(X_i)] = I(f) \tag{5.1}$$

(i.e. **find a function** g **such that** $g(x)f_X(x) = f(x)$ **for all** x) then will the arithmetic mean of the stochastic variables $\{g(X_i)\}_{i=1}^n$ converge to $I(f)$. Put in another way, the integral computation can be approximate by the sum of random variables.

For proper integrals, i.e., integrals over a bounded interval, the most straightforward approach is to simulate variables of uniform distribution.

Definition 5.1. *A stochastic variable, X, is uniform(a,b)-distributed if it has density function $f_X(x) = \frac{1}{b-a}$ for all $x \in [a,b]$ and zero elsewhere.*

For improper integrals, the uniform distribution is inadequate. But any distribution defined on the same set as the integral, with a corresponding g fulfilling condition (5.1), 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 finding a function g such that condition (5.1) holds is straightforward in the case of ordinary Monte Carlo integration since $g(x) = f(x)$ does the job.

5.3 Error and the central limit theorem

The LLN gives us the mean behavior when $n \to \infty$. Mathematically this may be 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 error, or spread, we use the *Central Limit Theorem* (CLT), telling us that the sample mean of a random variable with expected value μ and variance σ^2, is approximately normal $\mathrm{N}(\mu, \sigma^2/n)$-distributed.

Theorem 5.3.1 (The Central Limit Theorem)**.** *Let $X_1, X_2, ...$ be i.i.d. with $\mathbb{E}[X_i] = \mu$, $VAR(X_i) = \sigma^2 \in (0,\infty)$. 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.

If $\{x_i\}_{i=1}^n$ are observations of i.i.d. random variables, then the 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,

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

In particular, 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.

5.3.1 Examples

Example 5.1. Let $f(x) = \frac{4}{1+x^2}$, and employ ordinary Monte Carlo integration to compute the integral

$$I(f) = \int_0^1 f(x)dx = \int_0^1 \frac{4}{1+x^2}dx.$$

which thus is approximated with

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

where x_i are random numbers. One possibility is to take x_i to be (0,1)-uniformly distributed random numbers. A computer program for this could look as follows:

```
S_n=0, Errorterm=0
For 1 to n
    Generate a uniformly distributed random variable x_i.
    Calculate y=4/(1+x_i^2)
    S_n=S_n+y and Errorterm=y^2+Errorterm
End
S_n=S_n/n
Error=sqrt(Errorterm/n-S_n^2)/sqrt(n)
```

5.4 Variance reduction

Since variance of S_n is associated with the estimators performance, it is 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!

It should be noticed that the techniques often carry fancy names, but the ideas behind are relative straightforward.

5.4.1 Importance sampling

By choosing 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*. Thus, the uniform distribution is replaced by another distribution of sampling points.

First notice that

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

If we select p to be a probability density function, we may, as an alternative to ordinary Monte Carlo integration, generate random observations $x_1, \ldots, x_n$ with this probability density function, and approximate the integral I with

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

The error of this Monte Carlo approximation is $\sigma(f(X)/p(X))/\sqrt{(}n)$, where $\sigma^2(f(X)/p(X))$ is estimated as before, with

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

If the shape of p is similar 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 5.2. *Continued example. Instead of using ordinary Monte Carlo integration as in example 5.1, let $X_1, ..., X_n$ 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 $\int_{-\infty}^{\infty} f_X(x)dx = 1$. Furthermore we can see in Figure 5.1 that the shape of f_X is similar to f.*

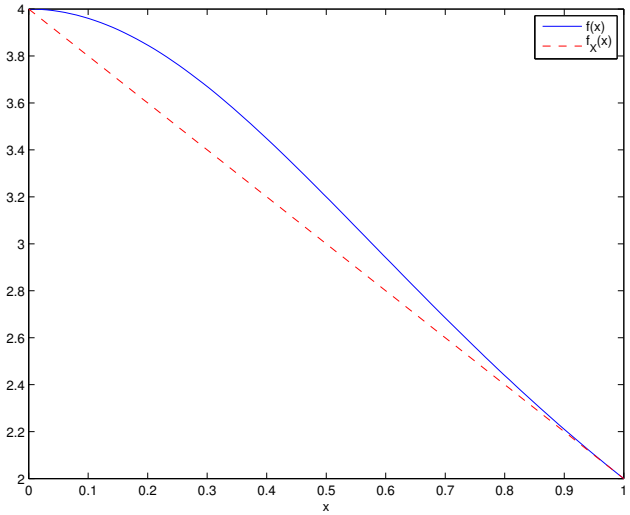

Figure 5.1: The functions $f(x)$ and (a scaled version of) $f_X(x)$.

Then the Monte Carlo approximation of $I(f)$ is

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

for observations $x_1, ..., x_n$ of the specific distribution. Since f and f_X have similar shape, f/f_X does not fluctuate much for $x \in [0,1]$ and thus

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

is small.

5.4.2 Control variates

One simple approach to reduce variance, is try to employ a *control variate* g, which is a function that is similar to f, and with a known value $I(g)$ of the integral. By linearity of integrals then

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

with g similar to f, the variance of $f - g$ should be smaller than that of f, and the integral $I = I(f)$ is approximated by the sum S_n of the Monte Carlo approximation of that integral and $I(g)$:

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

Note that this way the variation of S_n originates only from the difference between $f - g$.

Example 5.3. *Continued example. Instead of using importance sampling Monte Carlo integration in Example 5.2, if $g(x) = 4 - 2x$ for $x \in [0,1]$ and zero elsewhere then $I(g) = 3$. Monte Carlo integration of $I(f)$ employed with this control variate is then*

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

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

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

small.

5.4.3 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 and pairs of observations that are negatively correlated with each other are used to reduce the total variance (or error), based on the fact that:

$$\mathbf{Var}\{X + Y\} = \mathbf{Var} + \mathbf{Var}\{Y\} + 2\mathbf{Cov}\{X, Y\}.$$

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

Example 5.4. *Let $f : [0,1] \to \mathbb{R}$ 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 they are negatively correlated. If X is big then $1 - X$ has to be small. Further if f is monotone the non-positive correlation holds for $f(X)$ and $f(1 - X)$. In theory, any transformation, $T : [0,1] \to [0,1]$, is possible to employ to the ordinary Monte Carlo integration. Further if $f(X)$ and $f(T(X))$ are negatively correlated then the variance of

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

is less than the corresponding variance for ordinary Monte Carlo integration.

$1 - x_i$ is an equally good observation of a uniform(0,1) variable as x_i and x_i and $1 - x_i$ are negatively correlated. Thus $f(x_i)$ and $f(1 - x_i)$ are negatively correlated, since f is monotone. Thus the variance of S_n is reduced.

5.4.4 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 $M_1, \ldots, M_k$. 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 becomes

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

with the corresponding error

$$\Delta_{SS} = \sqrt{\sum_{j=1}^{k}\frac{\mathrm{vol}(M_j)^2}{n_j}\sigma^2_{M_j}(f)},$$

where

$$\sigma^2_{M_j}(f) = \left(\frac{1}{\mathrm{vol}(M_j)} \int_{M_j} f(x)^2 dx - \left(\frac{1}{\mathrm{vol}(M_j)} \int_{M_j} f(x) dx \right)^2 \right).$$

The variances $\sigma^2_{M_j}(f)$ of the different parts of the partition, in turn, are again estimated by means of Monte Carlo integration.

In order for stratified sampling to perform optimally, on should try to select

$$n_j \sim \mathrm{vol}(M_j)\sigma_{M_j}(f).$$

Example 5.5. *Let $f : [-1, 1] \to \mathbb{R}$ 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.By splitting the integration interval in these two parts, employing ordinary Monte Carlo integration on them separately, the technique of stratified sampling is utilized.*

5.5 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.

5.5.1 General theory for simulation of random variables

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

Lemma 5.1. *For a distribution function F, define the* generalized right-inverse *$F^{\leftarrow}$ by*

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

We have

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

Proof. [1]For $F(x) < y$ 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 $F^{\leftarrow}(y) \leq x \Leftrightarrow y \leq F(x)$. □

From a *random number*, i.e. a random variable that is *uniformly distributed over the interval* $[0, 1]$, a random variable with any other desired distribution can be simulated, at least in theory:

[1]This proof is not important for the understanding of the rest of the material.

Theorem 5.1. *If F is a distribution function and ξ a random number, 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 5.1 shows that

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

When using Theorem 5.1 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 5.1. *Let F be a continuous distribution function. Assume that there exists numbers $-\infty \le a < b \le \infty$ such that*

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

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

Corollary 5.1 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:

Example 5.6. The distribution function of an $\exp(\lambda)$-distribution with mean $1/\lambda$ $F(x) = 1 - e^{-\lambda x}$ for $x > 0$ has the inverse

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

Hence, if ξ is a random number, then Corollary 5.1 shows that

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

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

It is easy to simulate random variables with a discrete distribution:

Theorem 5.2 (Table Method)**.** *Let f be the probability density function for a discrete random variable with the possible value $\{y_1, y_2, y_3, \ldots\}$. If ξ is a random number, then the random variable*

$$\eta = \begin{cases} y_1 & \text{if} \quad 0 > \xi \leq f(y_1) \\ y_2 & \text{if} \quad f(y_1) < \xi \leq f(y_1) + f(y_2) \\ y_3 & \text{if} \quad f(y_1) + f(y_2) < \xi \leq f(y_1) + f(y_2) + f(y_3) \\ & \vdots \end{cases}$$

is a discrete random variable with the possible value $\{y_1, y_2, y_3, \ldots\}$ and probability density function $f_\eta = f$.

Proof. One sees directly that the result is true. Alternatively, the theorem can be shown by application of Theorem 5.1. □

5.5.2 Simulation of normal distributed random variables

It is possible to simulate Normal distributed random variables by an application of the Theorem 5.1. 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 5.3 (Box-Müller)**.** *If ξ and η are independent uniform(0,1) random variables, then we have*

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

Proof. [2]For 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 $\chi^2(2)$-distribution. Moreover, a $\chi^2(2)$-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 $[0, 2\pi]$.

Adding things up, and using Example 5.6, it follows that, for ξ and η independent uniform(0,1) random variables,

$$(N_1, N2) =_{\text{distribution}} \sqrt{-2\ln(\xi)}(\cos(2\pi\eta), \sin(2\pi\eta)). \quad \square$$

5.5.3 Simulation of γ-distributed random variables

As for normal random variables, by an application of Theorem 5.1 is it possible but intractable to simulate γ distributed random variables by the inverse of the distribution function. For integer values of the shape parameter, k, is a straightforward method of generating this data by summing k independent $\exp(\lambda)$ distributed variables.

[2]This proof is not important for the understanding of the rest of the material.

Theorem 5.4 (Erlang distribution)**.** *If* $\{\xi_i\}_{i=1}^k$ *are i.i.d.* $\exp(\lambda)$ *distributed random variables, then we have*

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

Proof. [3]Let $f_{\exp(\lambda)}$ 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

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

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

5.6 Software

The computer assignment is to be done in C. Since plotting figures is troublesome for non native C-programmers then use matlab or R 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.

```
/* 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
```

[3]This proof is not important for the understanding of the rest of the material.

```
    return;
}

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

Here gcc is the command that invokes the C compiler, -o is a flag that indicates the we want to have the compiled program in the file app. filename.c is the name of the file where the C code is stored. Further, -lm is a flag that have to be included when we use the math.h library.

The above asumes that a UNIX based system such as Linux or Mac is used. The easiest way to compile and execute C programs on the Windows computers in the computer rooms is to use an online compiler such as JDoodle, `https://www.jdoodle.com/c-online-compiler`. You should be able to solve this lab using it, or at the very least get started. It is however recommended to use Linux computers in the computer rooms for this lab.

5.7 Computer assignment

No report is needed on this lab. To pass the exercise you need to finish Assignment 1 and show the results to the exercise teacher or send the answers by email in pdf format to `statdata.chalmers@analys.urkund.se`. If you hand in by email, the pdf should contain your table, brief text answering the questions and the code you wrote to solve the lab. Answer the part marked in bold on Assignment 1. Assignment 2 and 3 are optional and can be done if you have time at the end of the lab .

Assignment 1 (Mandatory): It is well known that the number π can be calculated numerically as the integral

$$\pi = \int_0^1 \frac{4}{1+x^2} dx.$$

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 $n = 10^5, 10^6, 10^7, ...$ Perform an error estimate pretending that the real value of π is unknown and compare it with the actual error calculated using the real value of π. Begin with plotting the function $f(x) = 4/(1+x^2)$ to get a feeling for how it behaves (plot using Matlab or R).

- Implement the four variance reduction techniques presented here 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 (i.e. is the variance/error reduced)? 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.

  To hand in to the exercise teacher: Prepare a table that shows the Monte Carlo estimate of the integral, using the standard method and when implementing the four different variance reduction techniques. Present for different sample sizes n. Prepare another table showing the estimated error and the "real" error (since you know what the answer should be), using the different variance reduction techniques.

  For each method of variance reduction: Why is the variance reduced by the method? When is it more/less effective? Why?

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

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

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

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

Assignment 3 (Optional)

- In many applications, it is of interest to study worst case scenarios, and the *expected shortfall* $\mathbf{E}\{S_X(u)\}$ is a measure that is commonly used, for that purpose. Expected shortfall is defined as the expectation of a suitable *loss random variable* X, given that the loss is greater than a certain threshold u:

$$\mathbf{E}\{S_X(u)\} = \mathbf{E}\{X|X > u\}.$$

  Expected shortfalls can be difficult to calculate analytically, but with Monte Carlo simulations things simplify.

  Assume that an insurance company has found that the probability to have a flood is p, and that if a flood occurs, then the loss is exponential distributed with parameter λ. In other words, we have the loss $X = YZ$, where Y is a Bernoulli(p)-distributed random variable, and Z is an exp(λ)-distributed random variable with mean $1/\lambda$, independent of Y.

  Select $p = 0.1$, $1/\lambda = 3.4$ and $u = 10$, and use Monte Carlo simulation to estimate the expected shortfall $\mathbf{E}\{S_X(u)\}$. Also estimate the MC error as you did previously.

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