

Chapter 4

Monte Carlo Methods

So far we have learned how to simulate statistical models on a computer. In this section we will discuss how such simulations can be used to study properties of the underlying statistical model. We first address the approach to directly generate a large number of samples from the given model. The idea is then that the samples reflect the statistical behavior of the model; questions about the model can then be answered by studying statistical properties of the samples. The resulting methods are called *Monte Carlo* methods.

Since many interesting questions can be reduced to computing the expectation of some random variable, we will mostly restrict our attention to the problem of computing expectations of the form $\mathbb{E}[\Psi(X)]$, where X is a random sample from the system under consideration and $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ determines some output quantity of interest (QoI) in the system. There are several different methods to compute such an expectation:

- sometimes we can find the answer analytically. For example, if $X \sim f$, then

$$\mathbb{E}[\Psi(X)] = \int \Psi(x)f(x)dx;$$

- if the integral above cannot be solved analytically, we can try to use numerical quadrature formulas to get an approximation to the value of the integral. When X takes values in a low-dimensional space, this method often works well, but for higher dimensional spaces it can become very expensive and the resulting method may no longer be efficient;
- Monte Carlo estimation or Monte Carlo integration. This technique is based on the strong law of large numbers: if $\{X^{(j)}\}_{j \in \mathbb{N}}$ is a sequence of i.i.d. random variables with the same distribution as X , then with probability 1

$$\mathbb{E}[\Psi(X)] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \Psi(X^{(j)}). \quad (4.1)$$

Our aim for this section is to study approximations for $\mathbb{E}[\Psi(X)]$ based on (4.1).

Definition 4.0.1. A Monte Carlo method for estimating the expectation $\mathbb{E}[\Psi(X)]$ is a numerical method based on the approximation

$$\mathbb{E}[\Psi(X)] \approx \frac{1}{N} \sum_{j=1}^N \Psi(X^{(j)})$$

where $\{X^{(j)}\}_{j \in \mathbb{N}}$ are i.i.d. with the same distribution as X .

Since this is the goal
we have to learn to
approximate integrals
via Monte Carlo (since
expectations are integr.)

In order to compute a numerical value for this approximation we will need to generate a large number of samples X_j from the model: this can be done using the techniques from Sections 1 and 2 – in the present section we will assume that we have solved the problem of generating these samples and we will concentrate on the Monte Carlo estimate.

Example 4.0.1. For $\Psi(x) = x$, the Monte Carlo estimate reduces to

$$\mathbb{E}[X] \approx \frac{1}{N} \sum_{j=1}^N X^{(j)} = \bar{X}.$$

Example 4.0.2. Assume that we want to compute the expectation $\mathbb{E}[\sin(X)^2]$ where $X \sim N(\mu, \sigma^2)$. Generating N independent, (μ, σ^2) -distributed random variables $X^{(1)}, \dots, X^{(N)}$ we have

$$\mathbb{E}[\sin(X)^2] \approx \frac{1}{N} \sum_{j=1}^N \sin(X^{(j)})^2.$$

4.1 Properties of the Monte Carlo estimator

Let us consider a random variable Z that is the output quantity of interest (QoI) of a stochastic model, and the goal of computing its expectation:

$$\mu = \mathbb{E}[Z].$$

The Monte Carlo method consists simply of generating N i.i.d. replicas $Z^{(1)}, \dots, Z^{(N)}$ of Z and estimating $\mu = \mathbb{E}[Z]$ by a sample mean estimator,

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N Z^{(i)}.$$

e.g. before it was $Z = \Psi(X)$
 ~ notice that we're making no assumptions on the distribution of Z

We assume here that $\text{Var}[Z] = \sigma^2 < \infty$. Since the estimate $\hat{\mu}$ is constructed from random samples $Z^{(1)}, \dots, Z^{(N)}$, it is a random quantity itself. Some remarkable properties of the sample mean estimator are the following ones:

- $\hat{\mu}$ is unbiased, that is,

$$\mathbb{E}[\hat{\mu}] = \mu,$$

with the expectation taken with respect to the distribution of the sample $(Z^{(1)}, \dots, Z^{(N)})$ – if we repeat infinitely many times the experiment and we average the obtained results, we will obtain the exact mean μ .

- $\text{Var}[\hat{\mu}] = \sigma^2/N$. Indeed,

$$\begin{aligned} \text{Var}[\hat{\mu}] &= \mathbb{E}[(\hat{\mu} - \mathbb{E}[\hat{\mu}])^2] = \mathbb{E}\left[\left(\frac{1}{N} \sum_i (Z^{(i)} - \mu)\right)^2\right] = \frac{1}{N^2} \sum_{i,j} \mathbb{E}[(Z^{(i)} - \mu)(Z^{(j)} - \mu)] \\ &= \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}[(Z^{(i)} - \mu)^2] + \frac{1}{N^2} \sum_{i \neq j} \mathbb{E}[(Z^{(i)} - \mu)(Z^{(j)} - \mu)] = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N} \end{aligned}$$

since $\mathbb{E}[(Z^{(i)} - \mu)(Z^{(j)} - \mu)] = \sigma^2$ for all $i = 1, \dots, N$, and $\mathbb{E}[(Z^{(i)} - \mu)(Z^{(j)} - \mu)] = 0$ since $Z^{(i)}$ and $Z^{(j)}$ are independent.

- by the Strong Law of Large Numbers (SLLN),

$$\hat{\mu} \xrightarrow{a.s.} \mu \quad \text{as } N \rightarrow \infty;$$

(Strong) Law of Large Numbers. Assume that $\{Z^{(j)}\}$, $j = 1, 2, \dots$ are independent, identically distributed (i.i.d) with $\mathbb{E}[|Z^{(j)}|] < \infty$ and $\mathbb{E}[Z^{(j)}] = \mu$. Then $\hat{\mu} \xrightarrow{a.s.} \mu$ as $N \rightarrow \infty$, that is,

$$P\left(\left\{\frac{1}{N} \sum_{j=1}^N Z^{(j)} \rightarrow \mu\right\}\right) = 1.$$

- by the Central Limit Theorem (CLT), since $\text{Var}[Z] < \infty$,

$$\frac{\sqrt{N}(\hat{\mu} - \mu)}{\sigma} \xrightarrow{d} N(0, 1) \quad \text{as } N \rightarrow \infty,$$

i.e., the properly rescaled estimator $\frac{\sqrt{N}(\hat{\mu} - \mu)}{\sigma}$ converges in distribution to a standard normal random variable as $N \rightarrow \infty$.

Central Limit Theorem. Assume that $\{Z^{(j)}\}$, $j = 1, 2, \dots$ are independent, identically distributed (i.i.d) with $\mathbb{E}[Z^{(j)}] = \mu$, $\text{Var}[Z^{(j)}] = \sigma^2 < \infty$, and denote by $\hat{\mu} = \frac{1}{N} \sum_{j=1}^N Z^{(j)}$ the sample mean estimator. Then

$$\sum_{j=1}^N \frac{Z^{(j)} - \mu}{\sigma \sqrt{N}} = \frac{\sqrt{N}(\hat{\mu} - \mu)}{\sigma} \xrightarrow{d} \nu$$

where ν is $N(0, 1)$ and \xrightarrow{d} denotes convergence in distribution (or weak convergence).

Starting from the CLT, we can give an asymptotic $(1 - \alpha)$ -confidence interval

$$I_\alpha = \left[\hat{\mu} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}}, \hat{\mu} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}} \right]$$

once again: we obtain it without assumptions on the distribution of Z

with z_α the α -quantile of the normal distribution, i.e., $\Phi(z_\alpha) = \alpha$ and Φ the CDF of a standard normal random variable. Hence,

$$P(\mu \in I_\alpha) \xrightarrow{N \rightarrow \infty} 1 - \alpha;$$

equivalently, the error $|\mu - \hat{\mu}|$ satisfies

$$|\mu - \hat{\mu}| \leq z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}} \quad \text{with probability } 1 - \alpha \text{ asymptotically.}$$

The CLT shows that the Monte Carlo error is of the order $N^{-1/2}$, which is a very slow convergence rate (to reduce the error by a factor of 10, one has to multiply N by a factor of 100) and is characteristic of Monte Carlo estimates, generally not improvable. On the other hand, it holds under quite weak assumptions ($\text{Var}[Z] < +\infty$).

The previous error estimate is not practical, as it involves the – usually unknown – variance $\sigma^2 = \text{Var}[Z]$. We can replace this by the *sample variance estimator*

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (Z^{(i)} - \hat{\mu})^2;$$

Moreover, another good thing is that this does not depend on the dimension we are in (in whatever dimension we have that the order is $N^{-1/2}$)

this estimator is also unbiased and almost surely converging to σ^2 . It follows that $\sigma/\hat{\sigma} \rightarrow 1$ almost surely and

$$\frac{\sqrt{N}(\hat{\mu} - \mu)}{\hat{\sigma}} = \frac{\sigma}{\hat{\sigma}} \frac{\sqrt{N}(\hat{\mu} - \mu)}{\sigma} \xrightarrow{N(0, 1)} N(0, 1) \quad \text{as } N \rightarrow \infty,$$

thanks to the Slutsky theorem¹. A computable confidence interval is given by

$$\hat{I}_\alpha = \left[\hat{\mu} - z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{N}}, \hat{\mu} + z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{N}} \right]$$

fulfilling

$$P(\mu \in \hat{I}_\alpha) \xrightarrow{N \rightarrow \infty} 1 - \alpha.$$

it is possible to
improve this result
(however we need to
pay something more
to get something more)

4.1.1 Other asymptotic results

More generally speaking, we have that for any $z_0 > 0$, we can approximate

$$P\left(\left|\frac{\sqrt{N}(\mu - \hat{\mu})}{\sigma}\right| \leq z_0\right) \approx P(|\nu| \leq z_0) = 2\Phi(z_0) - 1$$

or, equivalently,

$$P\left(|\mu - \hat{\mu}| \leq z_0 \frac{\sigma}{\sqrt{N}}\right) \approx 2\Phi(z_0) - 1$$

being $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$. By the Berry-Esseen theorem, this bound can be further refined, and we find that, for any $z_0 > 0$,

$$P\left(|\mu - \hat{\mu}| \leq z_0 \frac{\sigma_Y}{\sqrt{M}}\right) \geq 2\Phi(z_0) - 1 - 2 \frac{C_{BE}\lambda^3}{(1+z_0)^3\sqrt{N}}.$$

Berry-Esseen Theorem. Assume that $\{Z^{(j)}\}$, $j = 1, 2, 3, \dots$ are independent, identically distributed (i.i.d) with $\mathbb{E}[Z^{(j)}] = \mu$, $\text{Var}[Z^{(j)}] = \sigma^2 < \infty$. Consider the scaled random variable

$$Z_N = \frac{\sqrt{N}}{\sigma} \left(\frac{1}{N} \sum_{j=1}^N Z^{(j)} - \mu \right)$$

with cumulative distribution function $F_{Z_N}(x) = P(Z_N \leq x)$, for any $x \in \mathbb{R}$. Assume that

$$\lambda = \frac{(\mathbb{E}[|Z^{(j)} - \mu|^3])^{1/3}}{\sigma} < \infty$$

then we have a uniform estimate in the central limit theorem,

$$|F_{Z_N}(x) - \Phi(x)| \leq \frac{C_{BE}\lambda^3}{(1+|x|)^3\sqrt{N}}$$

where $C_{BE} \approx 30.517\dots$ is a constant.

¹If $X_n \rightarrow X$ (that is, X_n converges to X in distribution, or weakly) and $Y_n \rightarrow a$, a constant, in probability, then $Y_n X_n \rightarrow aX$ and $X_n + Y_n \rightarrow X + a$ in distribution. Recall that almost sure convergence implies convergence in probability.

The quantity λ^3 is also denoted as *skewness* of the random variable $Z^{(j)}$ with mean μ and standard deviation σ . If we consider the scaled deviations

$$\xi^{(j)} = \frac{Z^{(j)} - \mu}{\sigma}$$

the following result also holds.

Law of the iterated logarithm. Assume that $\{\xi^{(j)}\}$, $j = 1, 2, 3, \dots$ are independent, identically distributed (i.i.d) with zero mean and unit variance, i.e., $\mathbb{E}[\xi^{(j)}] = 0$, $\mathbb{E}[(\xi^{(j)})^2] = 1$. Then, almost surely,

$$\limsup_{N \rightarrow \infty} \frac{|\sum_{j=1}^N \xi^{(j)}|}{\sqrt{2N \log \log N}} = 1.$$

According to the law of the iterated logarithm, if we use the Monte Carlo estimator to estimate $\mathbb{E}[Z]$, then the error decays almost surely at the rate $\sqrt{2N \log \log N}$ divided by N , that is,

$$\limsup_{N \rightarrow \infty} \left| \frac{1}{N} \sum_{j=1}^N Z^{(j)} - \mu \right| \leq \frac{\sigma}{\sqrt{N}} \sqrt{2 \log \log N}.$$

Under some stricter assumptions than before we gain $\sqrt{2 \log(\log(N))}$ to somehow balance the \sqrt{N} at the denominator

4.1.2 Practical aspects

As an output of a Monte Carlo simulation, one should *always* provide, besides the point estimate $\hat{\mu}$, also an estimate of the error, by e.g. the $(1 - \alpha)$ -asymptotic confidence interval \hat{I}_α .

In practice, one would like to choose N so as to achieve a prescribed tolerance, for instance by requiring that

$$N : |\hat{I}_\alpha| \leq tol \quad \leftarrow \hat{|I_\alpha|} = |2 \cdot z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{N}}| \leq tol$$

for a prescribed tolerance tol . This can be done by a *sequential procedure*:

1. Do a pilot run with \bar{N} replicas (\bar{N} not too big) to estimate $\hat{\sigma}_{\bar{N}}^2$;
2. Based on the previously estimated variance, fix

$$N = \frac{z_{1-\frac{\alpha}{2}}^2 \hat{\sigma}_{\bar{N}}^2}{tol^2};$$

3. Run Monte Carlo with N as at the previous point. Recompute $\hat{\sigma}_N^2$: if $\hat{\sigma}_N^2 > \hat{\sigma}_{\bar{N}}^2$ then set $\bar{N} = N$ and go back to point 2.

4.1.3 Vector-valued outputs

The previous considerations extend easily to the case of a vector valued output $\mathbf{Z} = (Z_1, \dots, Z_m)$ and the estimation of its expectation $\boldsymbol{\mu} = \mathbb{E}[\mathbf{Z}]$. In a Monte Carlo approach, we generate i.i.d. replicas $\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(N)}$, set

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^N \mathbf{Z}^{(i)}$$

and estimate the covariance matrix

$$\hat{C} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{Z}^{(i)} - \hat{\boldsymbol{\mu}})(\mathbf{Z}^{(i)} - \hat{\boldsymbol{\mu}})^\top.$$

Then, almost surely $\hat{\mu} \rightarrow \mu$, $\hat{C} \rightarrow C$ and

$$(\hat{\mu} - \mu)^\top \hat{C}^{-1} (\hat{\mu} - \mu) \xrightarrow{\text{a.s.}} \chi_m^2$$

a chi-squared distribution with m degrees of freedom. Hence, a computable $(1 - \alpha)$ -asymptotic confidence region is

$$\hat{I}_\alpha = \left\{ \mathbf{y} \in \mathbb{R}^m : (\hat{\mu} - \mathbf{y})^\top \hat{C}^{-1} (\hat{\mu} - \mathbf{y}) \leq \frac{\chi_{m,1-\alpha}^2}{N} \right\}$$

where $\chi_{m,1-\alpha}^2$ is the $(1 - \alpha)$ -quantile of the χ_m^2 distribution, so that

$$P(\mu \in \hat{I}_\alpha) \xrightarrow{N \rightarrow \infty} 1 - \alpha.$$

Once again we underly that this depends on N (number of simulations) and not on m (dimension)

Did you know? :-)

The name of the Monte Carlo method is inspired by... the gambling casinos at the city of Monte Carlo in Monaco. In its present form, the method is attributed to Fermi, Von Neumann, and Ulam, who developed it for the solution of problems related to neutron transport during the secret research at Los Alamos for the construction of the atomic bomb during world war II. There are, however, indications of previous uses of methods based on selections of random numbers. In particular the name of Lord Kelvin is mentioned for a paper which dates back to 1901, and Gosset (better known with the pseudonym Student) used experimental sampling to support his theoretical studies of statistical distributions. Fermi himself used already MC techniques in the 30's in connection with neutron transport.

4.2 Monte Carlo to compute integrals

Consider a simple stochastic model

$$Z = \psi(X_1, \dots, X_d)$$

where $\mathbf{X} = (X_1, \dots, X_d)$ is a random vector with joint PDF $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ and $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$. Then

$$\mathbb{E}[Z] = \int_{\mathbb{R}^d} \psi(x_1, \dots, x_d) f(x_1, \dots, x_d) dx_1 \dots dx_d = \int_{\mathbb{R}^d} \psi(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

A Monte Carlo approximation of $\mu = \mathbb{E}[Z]$ consists of:

- generating N i.i.d. replicas $\mathbf{X}^{(i)} \sim f$;
- computing $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N \psi(\mathbf{X}^{(i)}) \approx \int_{\mathbb{R}^d} \psi(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$.

Hence, the formula

$$\frac{1}{N} \sum_{i=1}^N \psi(\mathbf{X}^{(i)})$$

can be seen as a quadrature formula to approximate the integral $\int_{\mathbb{R}^d} \psi(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$.

Conversely, let us consider the problem of computing an integral (general)

$$I = \int_{\mathbb{R}^d} \psi(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}$$

where $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ is the integrand and $w : \mathbb{R}^d \rightarrow \mathbb{R}_+$ is a nonnegative weight such that $\int_{\mathbb{R}^d} w(\mathbf{x}) d\mathbf{x} = 1$. Then, we can estimate the integral by a Monte Carlo formula

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N \psi(\mathbf{X}^{(i)}) \quad \text{with } \mathbf{X}^{(i)} \stackrel{i.i.d.}{\sim} w$$

where $\mathbf{X}^{(i)}$ are the nodes and $1/N$ are the weights to assign to every node

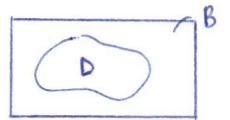
Hence we're exploiting MC for the evaluation of general integrals, even if we don't embed the story in a statistical setting

- Example 4.2.1. Let $D \subset \mathbb{R}^d$ be a bounded domain. We want to compute its volume $|D|$. Let B a rectangular domain containing D ; then

$$I = |D| = \int_D \mathbf{1}_D(\mathbf{x}) d\mathbf{x}$$

and

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_D(\mathbf{X}^{(i)}) = \frac{\#\{\mathbf{X}^{(i)} \in D\}}{N}, \quad \mathbf{X}^{(i)} \stackrel{i.i.d.}{\sim} \mathcal{U}(B),$$



i.e. we draw independently uniform points in B and count how many fall in D . The error in the Monte Carlo approximation is

$$|I - \hat{I}| \leq z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{N}}$$

with probability $1 - \alpha$ (asymptotically) where

$$\sigma^2 = \int_{\mathbb{R}^d} \psi^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} - I^2 \leq \int_{\mathbb{R}^d} \psi^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}.$$

Hence, the rate of convergence is $O(N^{-1/2})$ ad is achieved under the sole condition

$$\int_{\mathbb{R}^d} \psi^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} < +\infty$$

- that is, $\psi \in L_w^2(\mathbb{R}^d)$. Observe, in particular, that this rate is independent of the dimension d !
Although Monte Carlo has a very poor convergence rate $O(N^{-1/2})$, its use is still very appealing for high-dimensional problems.

L² weighted As a term of comparison, consider the problem of computing an integral

$$I = \int_{[0,1]^d} \psi(\mathbf{x}) d\mathbf{x}$$

let's consider for simplicity that $B = [0,1]^d$

on the unit hypercube using a tensor quadrature formula, e.g. the tensor mid-point rule,

$$I_{mp} = \sum_{i=1}^N \psi(\mathbf{x}^{(i)}) h^d$$

where $\mathbf{x}^{(i)}$ are the centers of the cells and $h = N^{-1/d}$ the cell length. The error of the formula can be bounded as²

$$|I - I_{mp}| \leq Ch^2 \|\psi\|_{C^2([0,1]^d)} = CN^{-2/d} \|\psi\|_{C^2([0,1]^d)}$$

therefore, such quadrature formula achieves a rate $N^{-2/d}$ (with respect to the number N of points used) provided $\psi \in C^2([0,1]^d)$ – hence, regularity is required in the integrand. For $d > 4$ the rate will be worse than Monte Carlo (this effect is usually called curse of dimensionality).

- curse of dimensionality

here we need to add regularity; this does not happen with Monte Carlo

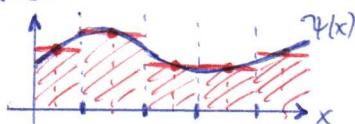
Hands-On Problem 4.1 focuses on the implementation in Matlab of the computation of integrals with Monte Carlo, and the comparison with a Gauss quadrature formula.

²More generally speaking,

$$|I - I_q| = O(n^{-r}) = O(N^{-r/d})$$

for higher-degree tensor quadrature rules which uses the same n point one-dimensional quadrature rule on each of d dimensions – and, as a result, $N = n^d$ function evaluations – for which r is the smaller of the number of continuous derivatives f has. Even modestly large d will give a bad result.

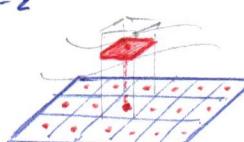
* d=1:



for every interval we evaluate the function $\psi(x)$ in the mid-point and the contribution of an interval to the integral will be:

$$(\text{length of the interval}) \cdot \psi(\text{mid point})$$

d=2



same thing as before: this time the mid point is the mid point of a little square, hence the contribution is: (area of the square) · $\psi(\text{mid point})$

- Example 4.2.2. Suppose we want to evaluate the integral

$$I = \int_a^b h(x)dx$$

for some function h where $x \in \mathcal{X} = \mathbb{R}$. Instead of considering classical quadrature formulas, let us write

$$I = \int_a^b h(x)dx = \int_a^b w(x)p(x)dx$$

where $w(x) = h(x)(b-a)$ and $p(x) = 1/(b-a)$ is the PDF of $\mathcal{U}([a,b])$. Hence,

$$I = \mathbb{E}[w(X)]$$

where $X \sim \mathcal{U}([a,b])$. By sampling $X^{(1)}, \dots, X^{(M)} \sim \mathcal{U}([a,b])$, by the law of large numbers we have

$$\hat{I} = \frac{1}{M} \sum_{m=1}^M w(X^{(m)}) \approx I$$

and the standard error of the estimate is

$$\hat{s}_e = \frac{s}{\sqrt{M}}, \quad s^2 = \frac{1}{M-1} \sum_{m=1}^M (h(X^{(m)}) - \hat{I})^2.$$

For example, suppose $h(x) = x^3$. Then $I = \int_0^1 x^3 dx = 1/4$. We can approximate this integral using Monte Carlo sampling as follows:

N
 $S = 10000;$
 $xs = \text{unifrnd}(0,1,S,1);$
 $samples = xs.^3;$
 $Ihat = \text{mean(samples)}$
 $se = \text{sqrt}(\text{var(samples)}/S)$

and obtain $I = 0.2525$ with standard error of 0.0028.

- Example 4.2.3. Suppose now we want to estimate π . We know that the area of a circle with radius r^2 is πr^2 . The area of a circle is given by

$$I = \int_{-r}^r \int_{-r}^r I(x^2 + y^2 \leq r^2) dx dy$$

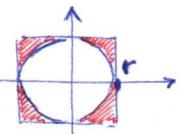
so that $\pi = I/r^2$. Let us approximate this by Monte Carlo integration. Let $X, Y \sim \mathcal{U}(-r, r)$, so using the above notation

$$w(x,y) = (b_x - a_x)(b_y - a_y)I(x^2 + y^2 \leq r^2) = (2r)(2r)I(x^2 + y^2 \leq r^2) = 4r^2 I(x^2 + y^2 \leq r^2), \quad p(x,y) = \frac{1}{4r^2}$$

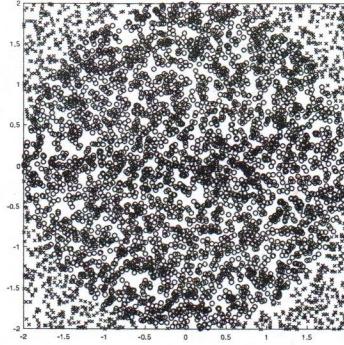
and we can implement this in Matlab as follows:

```
r=2; S=5000;
xs = unifrnd(-r,r,S,1); ys = unifrnd(-r,r,S,1);
rs = xs.^2 + ys.^2;
inside = (rs <= r.^2); evalution of w(x,y)
samples = 4*(r.^2)*inside;
Ihat = mean(samples); piHat = Ihat/(r.^2) because we evaluated the area of a circle of radius r
se = sqrt(var(samples)/S)
figure(1)
outside = ~inside;
plot(xs(inside), ys(inside), 'bo'); hold on
plot(xs(outside), ys(outside), 'rx'); axis square
```

We find $\hat{\pi} = 3.1416$ with standard error 0.09. Accepted/rejected points are reported in Figure 4.1.



we start sampling and we reject what falls in the red region

Figure 4.1: MC estimation of π

Hands-On Problem 4.2 focuses on the implementation in **Matlab** of the calculation of volumes with Monte Carlo.

- **Example 4.2.4.** Let us now assess the convergence of the MC method when computing the integral

$$I = \int_{[0,1]^d} f(x) dx, \quad f(x) : [0,1]^d \rightarrow \mathbb{R}.$$

Let $Y = f(\mathbf{X})$, where $\mathbf{X} \sim \mathcal{U}([0,1]^d)$. We have, denoting by $p(\mathbf{x})$ the uniform PDF,

$$I = \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} = \int_{[0,1]^d} f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \mathbb{E}[Y] \approx \frac{1}{N} \sum_{j=1}^N f(\mathbf{X}^{(j)}) = I_N$$

with $\mathbf{X}^{(j)} \sim \mathcal{U}([0,1]^d)$ i.i.d. The values $\mathbf{X}^{(j)}$ are sampled uniformly in the cube $[0,1]^d$, by sampling the components $X_i^{(j)}$ independently and uniformly on the interval $[0,1]$.

For instance, consider the computation of the integral

$$I = \int_{[0,1]^d} \frac{e^{\sum_{n=1}^d x_n}}{(e-1)^d} dx_1 \dots dx_d$$

through the following Matlab code:

```
N = 1e6; % Max. sample size
d = 20; % Dimension of the problem
u = rand(N,d); f = exp(sum(u'));
run_aver = cumsum(f)./(((1:N)').*(exp(1)-1)^d);
plot(1:N, run_aver),
figure, plot(1:N, run_aver), xlabel('N')
figure, plot(1:N, (run_aver-1)), xlabel('N')
figure, loglog(1:M,abs(run_aver-1)), xlabel('N') ,
hold on
s=1e-4
loglog(1:N,s./sqrt(1:N), 'k')
grid; xlabel('N')
```

the result is reported in Figure 4.2, left. Considering instead an ensemble of 20 iid runs we obtain the results reported in Figure 4.2, right.

Regarding the error, we have

— we want to see what happens when we vary N

the decay of the error is not monotone.
We compare the decay of the error with the trend of $1/\sqrt{N}$

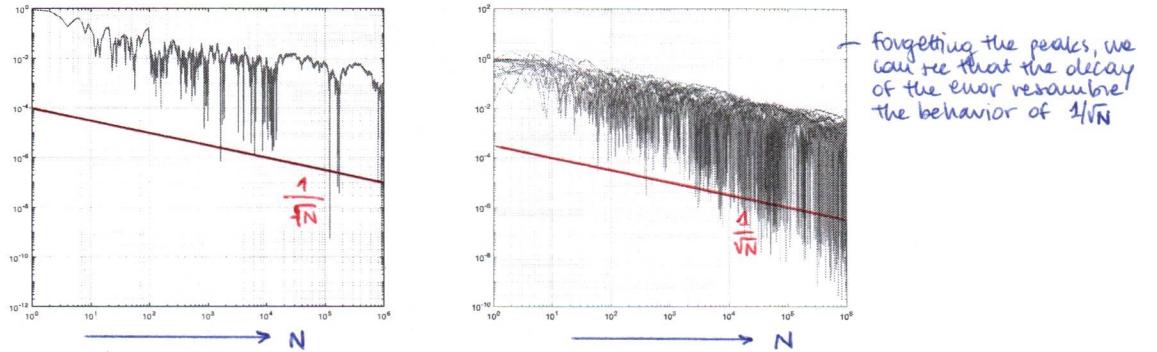


Figure 4.2: Computational error generated by Monte Carlo integration, single run (left) and 20 runs (right)

$$I_N - I = \frac{1}{N} \sum_{j=1}^N f(\mathbf{X}^{(j)}) - \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} = \sum_{j=1}^N \frac{f(\mathbf{X}^{(j)}) - \mathbb{E}[f(\mathbf{X})]}{N}$$

and, according to the CLT,

$$\sqrt{N}(I_N - I) \xrightarrow{\text{d}} \sigma \nu, \quad \text{where } \nu \text{ is } N(0, 1);$$

the exact variance

$$\sigma^2 = \int_{[0,1]^d} f^2(\mathbf{x}) d\mathbf{x} - \left(\int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} \right)^2 = \int_{[0,1]^d} \left(f(\mathbf{x}) - \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} \right)^2 d\mathbf{x}$$

is in practice approximated by the sample variance

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{j=1}^N \left(f(\mathbf{X}^{(j)}) - \sum_{j=1}^N \frac{f(\mathbf{X}^{(j)})}{N} \right)^2.$$

In Figure 4.3 we report the histograms based on 1000 MC outcomes, each based on 10 independent samples (e.g. from an exponential distribution with parameter $\lambda = 1/2$), as well as the CDF of a $N(0, 1)$ variable.

These results do not depend on the number of dimensions of the integral, which is the promised advantage of Monte Carlo integration against most deterministic methods that depend exponentially on the dimension. Notice also that, unlike in deterministic methods, the error estimate is not a strict error bound; random sampling may not uncover all the important features of the integrand that can result in an underestimate of the error.

4.3 Smooth functions of expectations and delta method

Consider several possible output variables $\mathbf{Z} = (Z_1, \dots, Z_m)$ of a stochastic model. In certain cases, the goal of the analysis is to compute a function of the expectation of \mathbf{Z} ,

$$\zeta = f(\mathbb{E}[Z_1], \dots, \mathbb{E}[Z_m])$$

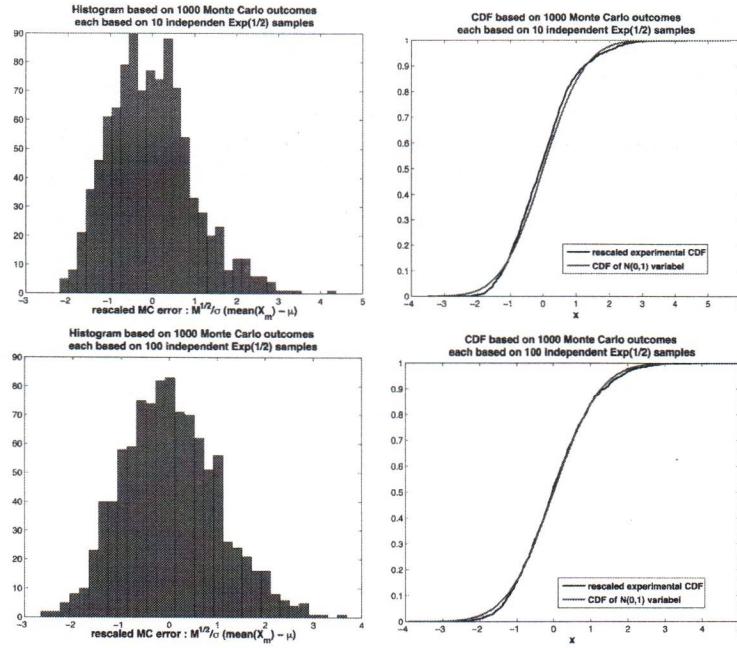


Figure 4.3: Histograms and CDF based on 1000 Monte Carlo outcomes

we want to be able to compute derivatives

with $f : \mathbb{R}^m \rightarrow \mathbb{R}$ smooth. Let $\mu_i = \mathbb{E}[Z_i]$ and $\mu = (\mu_1, \dots, \mu_m) \in \mathbb{R}^m$. The natural Monte Carlo estimator for ζ is

$$\hat{\zeta} = f(\hat{\mu}_1, \dots, \hat{\mu}_m), \quad \text{with } \hat{\mu}_i = \frac{1}{N} \sum_{k=1}^N Z_i^{(k)}$$

with $\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(N)}$ i.i.d. replicas of \mathbf{Z} . If f is continuous at μ , then $\hat{\zeta} \xrightarrow{a.s.} \zeta$, that is, $\hat{\zeta}$ is a consistent estimator of ζ .

- **Example 4.3.1.** Let Z be a random variable, output of a stochastic model. Suppose we want to estimate the coefficient of variation of Z ,

$$\zeta = \frac{\sigma(Z)}{\mu(Z)} = \sqrt{\frac{\mathbb{E}[Z^2] - \mathbb{E}[Z]^2}{\mathbb{E}[Z]}}.$$

Setting $\mathbf{Z} = (Z_1, Z_2) = (Z, Z^2)$, then we have

$$\zeta = f(\mathbb{E}[\mathbf{Z}]) = f(\mathbb{E}[Z_1], \mathbb{E}[Z_2]) = \sqrt{\frac{\mathbb{E}[Z_2]}{\mathbb{E}[Z_1]^2} - 1}.$$

The question is how to estimate the error on $\hat{\zeta}$, and provide a confidence interval. This can be done by the *delta method*, based on a first-order Taylor expansion around μ :

$$\hat{\zeta} - \zeta = f(\hat{\mu}) - f(\mu) = \nabla f(\mu) \cdot (\hat{\mu} - \mu) + o(\|\hat{\mu} - \mu\|).$$

Let $C = \text{Cov}(\mathbf{Z}) = \mathbb{E}[(\mathbf{Z} - \mu)(\mathbf{Z} - \mu)^\top]$. Then³

$$\sqrt{N}(\hat{\zeta} - \zeta) \xrightarrow{D} N(0, \nabla f(\mu) C \nabla f(\mu)). \quad \text{→ another version of the CLT.}$$

Notice that: $\nabla f(\hat{\mu}) \cdot C \cdot \nabla f(\hat{\mu}) \in \mathbb{R}$

³Here the gradient is considered as a row vector.

A computable confidence interval can then be constructed by replacing C with

$$\hat{C} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{Z}^{(i)} - \hat{\mu})(\mathbf{Z}^{(i)} - \hat{\mu})^\top$$

and $\nabla f(\mu)$ by $\nabla f(\hat{\mu})$. Observe, however, that the estimator $\hat{\zeta}$ is biased in general.

Hands-On Problem 4.3 focuses on a MC analysis on a very simple physical model.

$$|\mu - \hat{\mu}| \leq z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{N}}$$

- we cannot reduce $\frac{1}{\sqrt{N}}$ (this is the convergence rate)

1. - how can we reduce $\hat{\sigma}$?

→ variance reduction techniques (Chp. 5)

2. - how to relax the MC construction?

→ choices of points according to suitable rules that will allow us to get a convergence that will be

$$\frac{\log(N)}{\sqrt{N}}$$

here the dimension d may appear
(price to pay for having $\log(N)$)

(QUASI-MONTE CARLO formulas) (Chp. 6)

Here points won't be random. They'll be build with deterministic patterns.

Moreover, there's a way to combine different MC models. (Chp. 7)

→ Multi-fidelity MC
(/ Multi-level MC)