

**Monte Carlo Methods**  
Lecture notes for MAP001169  
Based on Script by Martin Sköld

adopted by Krzysztof Podgórski



# Contents

<b>I</b>	<b>Simulation and Monte-Carlo Integration</b>	<b>5</b>
<b>1</b>	<b>Simulation and Monte-Carlo integration</b>	<b>7</b>
1.1	Issues in simulation . . . . .	7
1.2	Raw ingredients . . . . .	7
<b>2</b>	<b>Simulating from specified distributions</b>	<b>9</b>
2.1	Transforming uniforms . . . . .	9
2.2	Transformation methods . . . . .	9
2.3	Rejection sampling . . . . .	9
2.4	Conditional methods . . . . .	9
<b>3</b>	<b>Monte-Carlo integration</b>	<b>11</b>
3.1	Generic Monte Carlo integration . . . . .	11
3.2	Bias and the Delta method . . . . .	11
3.3	Variance reduction by rejection sampling . . . . .	11
3.4	Variance reduction by importance sampling . . . . .	12
3.4.1	Unknown constant of proportionality . . . . .	15



**Part I**

**Simulation and Monte-Carlo  
Integration**



## Chapter 1

# Simulation and Monte-Carlo integration

1.1 Issues in simulation

1.2 Raw ingredients





## Chapter 2

# Simulating from specified distributions

- 2.1 Transforming uniforms
- 2.2 Transformation methods
- 2.3 Rejection sampling
- 2.4 Conditional methods



## Chapter 3

# Monte-Carlo integration

### 3.1 Generic Monte Carlo integration

### 3.2 Bias and the Delta method

### 3.3 Variance reduction by rejection sampling

The method based on uniform sampling is simple but also not very intelligent. After all uniform distribution contains no information about the function at integral of which we aim. Through uniform samples we sometimes sample over regions where values of the function that do contribute much to the value of the integral but equally often (uniformly) we sample over regions where this is not true. Due to this we have large variability in the approximations – large variance. One can try to reduce this variability by being smarter, i.e. by utilizing some information about the function. In fact one method of achieving it can utilize rejection sampling algorithm that was discussed as a method of sampling from a distribution. There we were approximating a shape of the density (up to the normalizing constant) by the shape of a proposal density from which we could sample. The fact that the method worked without necessity of knowing the normalizing constants can be utilized here.

Consider a known density  $f(x)$  on  $(a, b)$  from which one can simulate samples. Let us assume that  $\phi(x) > 0$  is a function on  $(a, b)$  from which an integral is supposed to be approximated. Let a constant  $K$  be such that  $\phi(x) \leq Kf(x)$ . Consider 0-1 random variables  $X_i$  indicating if the  $i$ th attempt in rejection sampling is rejected or accepted. Then

$$P(X_i = 1) = \frac{\int_a^b \phi(x) dx}{K}.$$

Since  $X_i$  are iid thus by the LLN

$$\hat{I}_n = K\bar{X} \approx \int_a^b \phi(x) dx.$$

One can easily see that the variance of the method is given by

$$\text{Var}(\hat{I}_n) = K^2 \frac{\int_a^b \phi(x) dx}{K} \left( 1 - \frac{\int_a^b \phi(x) dx}{K} \right) / n = \int_a^b \phi(x) dx \left( K - \int_a^b \phi(x) dx \right) / n.$$

Thus if  $K$  is close to  $\int_a^b \phi(x) dx$  the variance can be small.

**Exercise 3.1.** Consider a distribution on  $[0, \pi]$  given by the cdf

$$F(u) = \frac{1 - e^{-u^2}/2}{1 - e^{-\pi^2}/2}.$$

The simulation from this distribution is easily achieved by inverting the cdf. One can use the discussed method to evaluate the integral of

$$\phi(x) = x\sqrt{\sin x}e^{-x^2}, \quad x \in [0, \pi].$$

Perform the analysis comparing variance of the method with the variance of the method based on uniform sampling over the interval  $[0, \pi]$ .

The idea of variance reduction that is evident in our rejection algorithm is further explored by a similar but slightly more advanced and more popular method that is discussed next.

### 3.4 Variance reduction by importance sampling

Importance sampling is a technique that might substantially decrease the variance of the Monte-Carlo error. It can also be used as a tool for estimating  $E(\phi(X))$  in cases where  $X$  can not be sampled easily.

We want to calculate

$$\tau = \int \phi(x)f(x)dx \tag{3.1}$$

which can be re-written

$$\tau = \int \psi(x)g(x)dx \tag{3.2}$$

where  $\psi(x) = \phi(x)f(x)/g(x)$ . Hence, if we obtain a sample  $x_1, x_2, \dots, x_n$  from the distribution of  $g$ , then we can estimate the integral by the unbiased estimator

$$t_n = n^{-1} \sum_{i=1}^n \psi(x_i), \tag{3.3}$$

for which the variance is

$$\text{Var}(T_n) = n^{-1} \int \{\psi(x) - \tau\}^2 g(x) dx. \tag{3.4}$$

This variance can be very low, much lower than the variance of an estimate based on draws from  $f$ , provided  $g$  can be chosen so as to make  $\psi$  nearly constant. Essentially what is happening is that the simulations are being concentrated in the areas where there is greatest variation in the integrand, so that the informativeness of each simulated value is greatest. Another important advantage of importance sampling comes in problems where drawing from  $f$  is difficult. Here draws from  $f$  can be replaced by draws from an almost arbitrary density  $g$  (though it is essential that  $\phi f/g$  remain bounded).

This example illustrates the idea. Suppose we want to estimate the probability  $P(X > 2)$ , where  $X$  follows a Cauchy distribution with density function

$$f(x) = \frac{1}{\pi(1+x^2)} \quad (3.5)$$

so we require the integral

$$\int \mathbf{1}\{x > 2\}f(x)dx. \quad (3.6)$$

We could simulate from the Cauchy directly and apply basic Monte-Carlo integration, but the variance of this estimator is substantial. As with the bivariate Normal example, the estimator is the empirical proportion of exceedances; exceedances are rare, so the variance is large compared to its mean. Put differently, we are spending most of our simulation budget on estimating the integral of  $\mathbf{1}\{x > 2\}f(x)$  over an area (i.e. around the origin) where we know it equals zero.

Alternatively, we observe that for large  $x$ ,  $f(x)$  is similar in behaviour to the density  $g(x) = 2/x^2$  on  $x > 2$ . By inversion, we can simulate from  $g$  by letting  $x_i = 2/u_i$  where  $u_i \sim U[0, 1]$ . Thus, our estimator becomes:

$$t_n = n^{-1} \sum_{i=1}^n \frac{x_i^2}{2\pi(1+x_i^2)}, \quad (3.7)$$

where  $x_i = 2/u_i$ . Implementing this with the R-function

```
impsamp=function(n){
  x=2/runif(n)
  psi=x^2/(2*pi*(1+x^2))
  tn=mean(psi)
  cum=cumsum(psi)/seq(1,n,by=1)
  impsamp=list(tn=tn,cum=cum)
}
```

and processing

```
is=impsamp(1000);
plot(is$cum, type='l')\
```

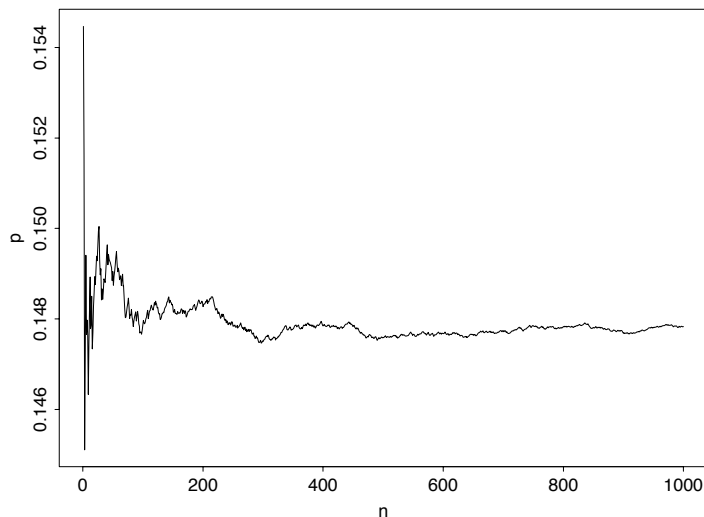


Figure 3.1: Convergence of importance sampled mean

gave the estimate  $t_n = .1478$ . The exact value is  $.5 - \pi^{-1} \tan 2 = .1476$ . In Figure 3.1 the convergence of this sample mean to the true value is demonstrated as a function of  $n$  by plotting the additional output vector `cum`.

For comparison, in Figure 3.2, we show how this compares with a sequence of estimators based on the sample mean when simulating directly from a Cauchy distribution. Clearly, the reduction in variability is substantial (the importance sampled estimator looks like a straight line).

### 3.4.1 Unknown constant of proportionality

To be able to use the above importance sampling techniques, we need to know  $f(x)$  explicitly. Just knowing  $Mf$  for an unknown constant of proportionality  $M$  is not sufficient. However, importance sampling can also be used to approximate  $M$ . Note that,

$$M = \int Mf(x) dx = \int \frac{Mf(x)}{g_2(x)} g_2(x) dx, \quad (3.8)$$

for a density  $g_2$ . Thus, based on a sample  $x_1, x_2, \dots, x_N$  from  $g_2$ , we can approximate  $M$  by

$$t_N = \frac{1}{N} \sum_{i=1}^N \frac{Mf(x_i)}{g_2(x_i)}. \quad (3.9)$$

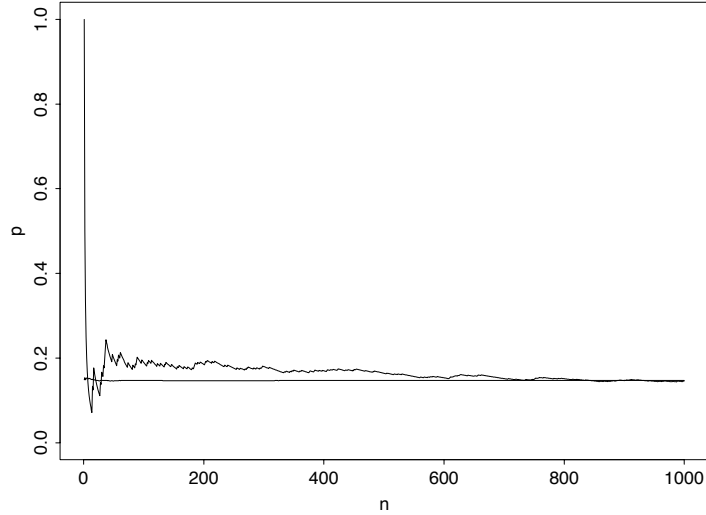


Figure 3.2: Comparison of importance sampled mean with standard estimator

It should be noted that this approximation puts some restrictions on the choice of  $g_2$ . To have a finite variance, we need (with  $X' \sim g_2$ )

$$E\left(\frac{(Mf(X'))^2}{g_2(X')^2}\right) = \int \frac{(Mf(x))^2}{g_2(x)} dx,$$

to be finite, i.e.  $f^2/g_2$  is integrable. Hence, a natural requirement is that  $f/g_2$  is bounded. This can now be used to approximate  $\tau = E(\phi(X))$  using sequences  $x_1, x_2, \dots, x_N$  from  $g_2$  and  $y_1, y_2, \dots, y_N$  from  $g$  through

$$\frac{t'_N}{t_N} = \left( \sum_{i=1}^N \frac{\phi(y_i) Mf(y_i)}{g(y_i)} \right) / \left( \sum_{i=1}^N \frac{Mf(x_i)}{g_2(x_i)} \right), \quad (3.10)$$

where the numerator approximates  $M\tau$  and denominator  $M$ . Of course we could use  $g = g_2$  and  $x_i = y_i$  in (3.10), but this is not usually the most efficient choice.