# R Programming for Simulation and MC Methods

## Numerical Integration

# Numerical Integration

It is frequently necessary to compute definite integrals $\int_a^b f(x)dx$ of a given function $f$. From the Fundamental Theorem of Calculus we know that if we can find an antiderivative or indefinite integral $F$, such that $F'(x) = \frac{d}{dx}F(x) = f(x)$, then $\int_a^b f(x)dx = F(b) - F(a)$. However for many functions $f$ it is impossible to write down an antiderivative in closed form. That is, we have no finite formula for $F$. In such cases we can use numerical integration to approximate the definite integral.

For example, in statistics we often use definite integrals of the standard normal density, that is, integrals of the form

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

We know that $\Phi(0) = 1/2$ and $\Phi(\infty) = 1$, but for all other $z$ numerical integration is used.

# Numerical Integration

In this chapter we consider three numerical integration techniques: the trapezoid rule, Simpson's rule, and adaptive quadrature. In each case we suppose that we are given an integrable[1] function $f(x)$ and an interval $[a, b]$ and the object is to approximate

$$\int_a^b f(x)\, dx.$$

We subdivide the interval $[a, b]$ into $n$ equal subintervals each of length $h = (b - a)/n$. The endpoints of these subintervals are labelled

$$a = x_0, x_1, x_2, \ldots, x_{n-1}, x_n = b.$$

We approximate the integral on each of these small intervals, then add all the small approximations to give a total approximation to the original integral.

# Trapezoidal Rule

## 11.1 Trapezoidal rule

The trapezoidal approximation is obtained by approximating the area under $y = f(x)$ over the subinterval $[x_i, x_{i+1}]$ by a trapezoid. That is, the function $f(x)$ is approximated by a straight line over the subinterval $[x_i, x_{i+1}]$ (Figure 11.1). The width of the trapezoid is $h$, the left side of the trapezoid has height $f(x_i)$ and the right side has height $f(x_{i+1})$. The area of the trapezoid is thus

$$\frac{h}{2}(f(x_i) + f(x_{i+1})).$$

Now we add the areas for all of the subintervals together to get our trapezoidal approximation $T$ to the integral $\int_a^b f(x)dx$:

**Trapezoidal rule**

$$T = \frac{h}{2}(f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)).$$

# Trapezoidal Rule

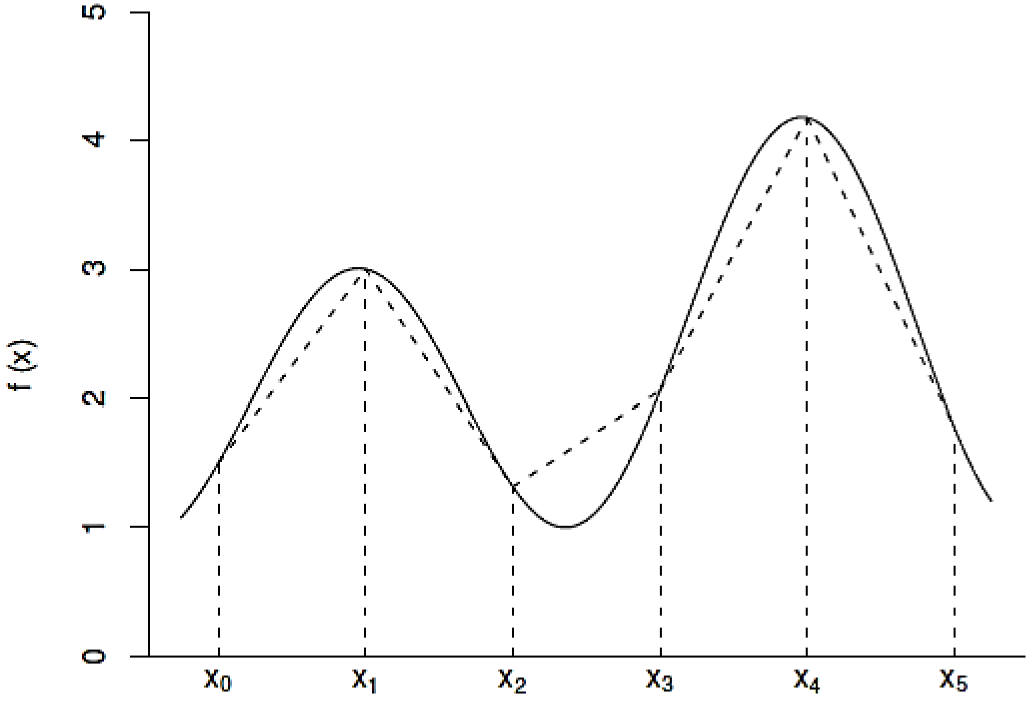


Figure 11.1 *The approximation of f used by the trapezoidal rule.*

# Trapezoidal Rule

**Trapezoidal rule**

$$T = \frac{h}{2}(f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)).$$

Notice that for $i = 1, \ldots, n-1$, $f(x_i)$ contributes to the area of the trapezoid to the left of $x_i$ and to the right of $x_i$ and so appears multiplied by 2 in the formula above. In contrast $f(x_0)$ and $f(x_n)$ contribute only to the area of the first and last trapezoid, respectively.

Here is an implementation in R. We use it to estimate $\int_0^1 4x^3\,dx = 1$.

# Simpson's Rule

## 11.2 Simpson's rule

Simpon's rule subdivides the interval $[a, b]$ into $n$ subintervals, where $n$ is even, then on each consecutive pair of subintervals, it approximates the behaviour of $f(x)$ by a parabola (polynomial of degree 2) rather than by the straight lines used in the trapezoidal rule.

Let $u < v < w$ be any three points distance $h$ apart. For $x \in [u, w]$ we want to approximate $f(x)$ by a parabola which passes through the points $(u, f(u))$, $(v, f(v))$, and $(w, f(w))$. There is exactly one such parabola $p(x)$ and it is given by the formula

$$p(x) = f(u)\frac{(x-v)(x-w)}{(u-v)(u-w)} + f(v)\frac{(x-u)(x-w)}{(v-u)(v-w)} + f(w)\frac{(x-u)(x-v)}{(w-u)(w-v)}.$$

# Simpson's Rule

As an approximation to the area under the curve $y = f(x)$, we use $\int_u^w p(x)dx$. A rather lengthy but elementary calculation shows

$$\int_u^w p(x)dx = \frac{h}{3}(f(u) + 4f(v) + f(w)).$$

Now, assuming that $n$ is even, we add up the approximations for the subintervals $[x_{2i}, x_{2i+2}]$ to obtain Simpson's approximation $S$ to the integral $\int_a^b f(x)dx$.

**Simpson's rule**

$$S = \frac{h}{3}(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots + 4f(x_{n-1}) + f(x_n)).$$

# Simpson's Rule

**Simpson's rule**

$$S = \frac{h}{3}(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots + 4f(x_{n-1}) + f(x_n)).$$

Notice that the $f(x_i)$ for $i$ odd are all weighted 4, while the $f(x_i)$ for $i$ even (except 0 and $n$) are weighted 2 as they each appear in two subintervals.

Obviously Simpson's rule gives exact results if $f(x)$ is a quadratic function since it is based on approximating each piece of $f(x)$ by a parabola. Surprisingly, it also gives exact results if $f(x)$ is a cubic function. In general it gives better results than the trapezoid rule.

# Example: Phi.r

### *11.2.1 Example:* $\Phi(z)$ `Phi.r`

One of Gauss' many prodigious acts was to compile by hand tables of $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$, estimated to several decimal places. (This is the distribution function of a normal or Gaussian random variable; see Section 16.5.1.) Thankfully we can now do this using a computer, as follows.

Running the command `source("../scripts/Phi.r")` we get the output given in Figure 11.2. We will see in Section 16.1 that R actually has a built-in function for calculating $\Phi(z)$, namely `pnorm`.

# Example: Phi.r

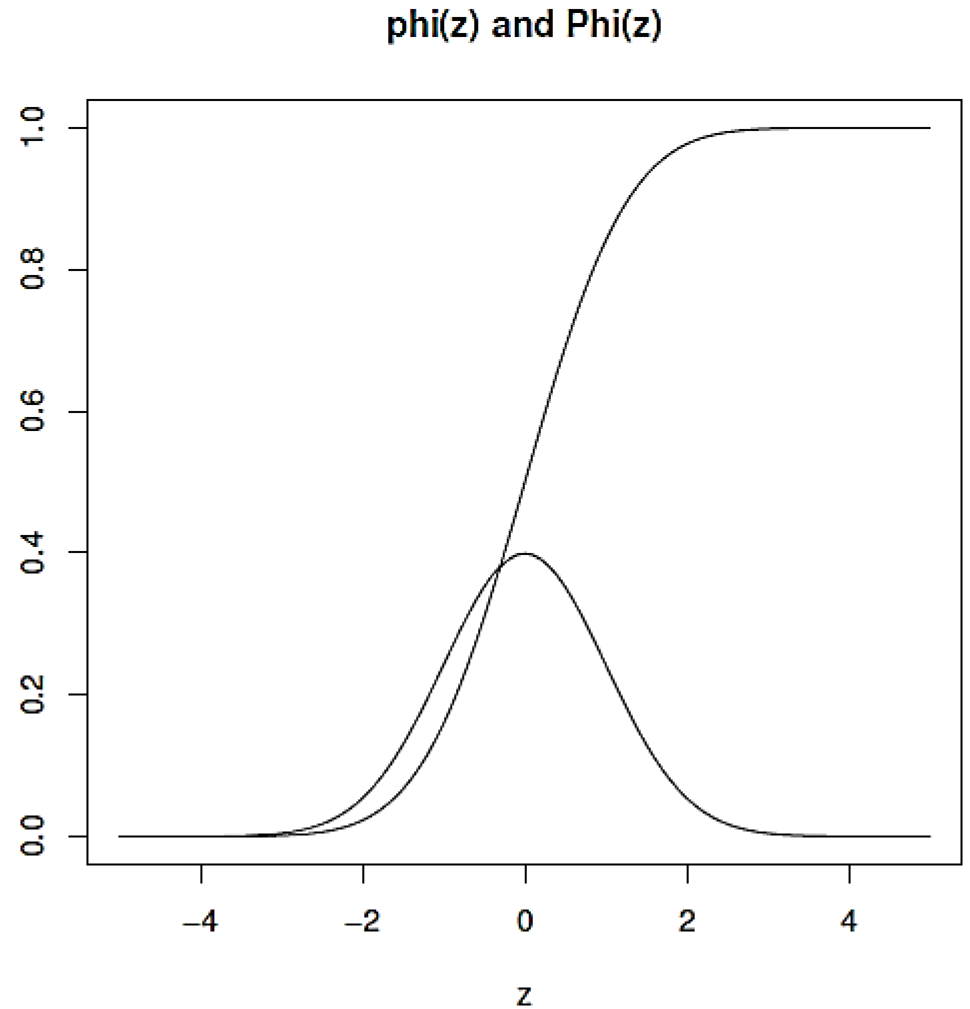


Figure 11.2 $\phi(z) = e^{-x^2/2}/\sqrt{2\pi}$ *and its integral* $\Phi$*; see Example 11.2.1.*

# Example: Phi.r

To test the accuracy of Simpson's rule we estimated $\int_{0.01}^{1}(1/x)\,dx = -\log(0.01)$ for a sequence of increasing values of $n$, the number of partitions. A plot of log(error) against $\log(n)$ appears to have a slope of roughly $-4$ for large values of $n$, indicating that the error decays like $n^{-4}$. This can in fact be shown to hold in general for functions $f$ with a continuous fourth derivative.

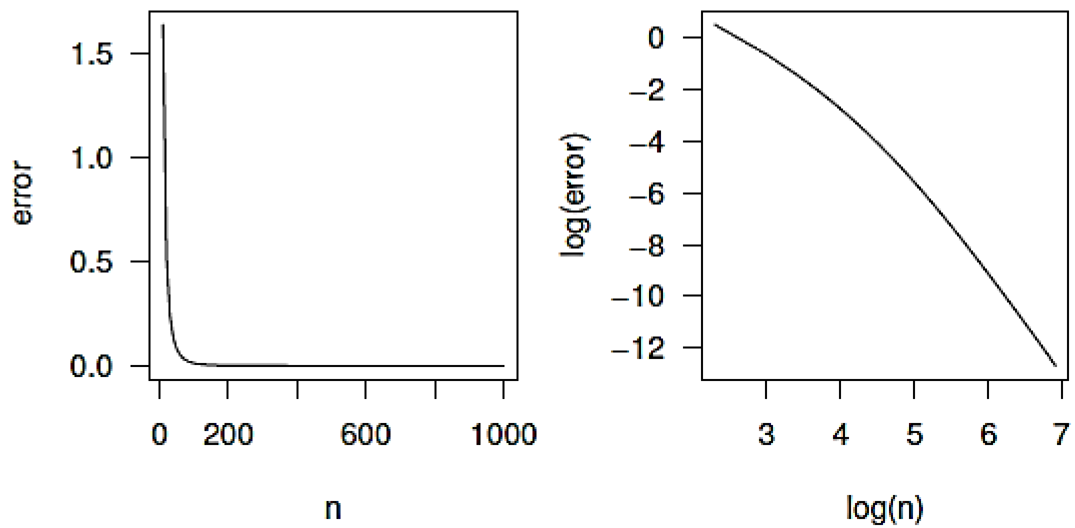


Figure 11.3 *Error using Simpson's method with a partition of size n; see Example 11.2.2.*

# Hit-and-Miss Method

We wish to calculate $I = \int_a^b f(x)dx$.

Let $c$ and $d$ be such that $f(x) \in [c, d]$ for all $x \in [a, b]$. Let $A$ be the set bounded above by the curve and by the box $[a, b] \times [c, d]$, then $I = |A| + c(b - a)$. Thus if we can estimate $|A|$ then we can estimate $I$. $A$ is illustrated as the shaded region in Figure 19.1.

To estimate $|A|$ imagine throwing darts at the box $[a, b] \times [c, d]$. On average the proportion that land under the curve will be given by the area of $A$ over the area of the box, that is by $|A|/((b - a)(d - c))$, giving us a means of estimating $|A|$.

# Hit-and-Miss Method

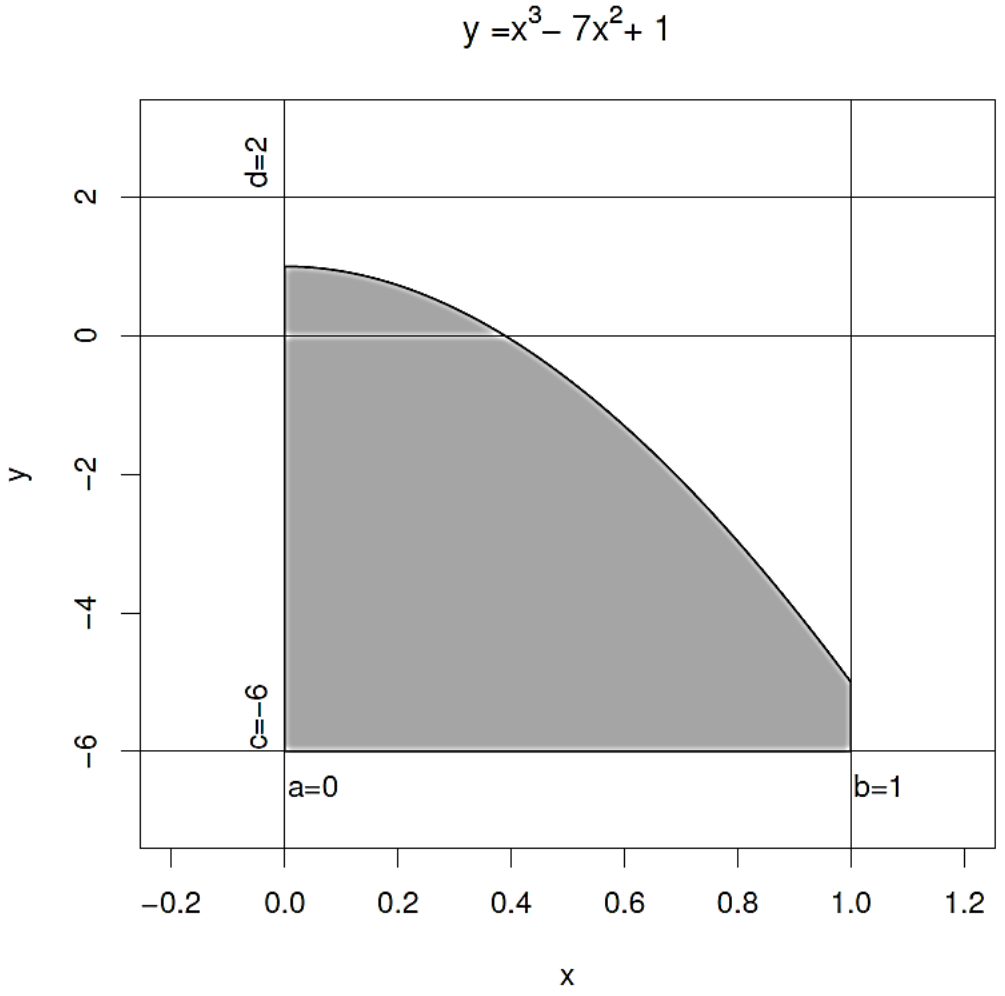


Figure 19.1 *The area of interest in the hit-and-miss method.*

# Hit-and-Miss Method

We apply the method to estimate

$$\int_0^1 (x^3 - 7x^2 + 1)dx = (x^4/4 - 7x^3/3 + x)\Big|_0^1$$
$$= -13/12 = -1.0833 \text{ (to 4 decimal places).}$$

Taking the min and max of each term we see that on $[0, 1]$ the function is bounded below by $c = 0 - 7 + 1 = -6$ and above by $d = 1 + 0 + 1 = 2$.

# Accuracy in Higher Dimensions

The big-O notation is used to describe how fast a function grows. We say $f(x)$ is $O(x^{-\alpha})$ if $\limsup_{x\to\infty} f(x)/x^{-\alpha} = \limsup_{x\to\infty} f(x)x^{\alpha} < \infty$.

Let $d$ be the dimension of our integral and $n$ the number of function calls used, then the accuracy of the different numerical integration techniques we have seen is as follows:

| Method | Error |
|---|---|
| Trapezoid | $O(n^{-2/d})$ |
| Simpson's rule | $O(n^{-4/d})$ |
| Hit-and-miss Monte-Carlo | $O(n^{-1/2})$ |
| Improved Monte-Carlo | $O(n^{-1/2})$ |

We see that the size of the error for the Monte-Carlo methods does not depend on $d$ and that, asymptotically, they are preferable when $d > 8$.