Approximation

Liming Pang

We have learned some techniques of integration during the previous several sections, and now we are able to deal with some complicated integrals. But there are more integrals that are hard to compute or impossible to express the accurate solution. In such circumstances, the best we would expect is to obtain a good estimation of the integral, and applications of mathematics to other subjects or the industry usually do not require an accurate value, so an efficient approximation of integrals will work in those cases.

1 Endpoint Approximation

A most natural trial for approximating definite integrals is to come back to its definition: Riemann Sum. Recall that a definite integral is the limit of the corresponding Riemann Sum

$$\int_a^b f(x)\,dx = \lim_{\max \Delta x_i \to 0} f(x_i^*)\Delta x_i$$

where $\Delta x_i = x_i - x_{i-1}$ and $x_i^* \in [x_{i-1}, x_i]$.

Recall that in practice, we have the freedom for choosing x_i and choosing $x_i^* \in [x_{i-1}, x_i]$. Different choices of x_i^* will lead to different ways of approximation.

We now let $x_i = a + \frac{b-a}{n}i$, then $\Delta x_1 = \Delta x_2 = ... = \Delta x_n = \Delta x = \frac{b-a}{n}$.

If we choose $x_i^* = x_{i-1}$, then the Riemann Sum is

$$L_n = (\sum_{i=0}^{n-1} f(x_i))\Delta x = (\sum_{i=0}^{n-1} f(x_i))\frac{b-a}{n}$$

So we get **Left Endpoint Approximation**:

$$\int_a^b f(x)\,dx \approx (\sum_{i=0}^{n-1} f(x_i))\Delta x$$

Similarly, if we choose $x_i^* = x_{i-}$, then the Riemann Sum is

$$R_n = (\sum_{i=1}^{n} f(x_i))\Delta x = (\sum_{i=1}^{n} f(x_i))\frac{b-a}{n}$$

So we get **Right Endpoint Approximation**:

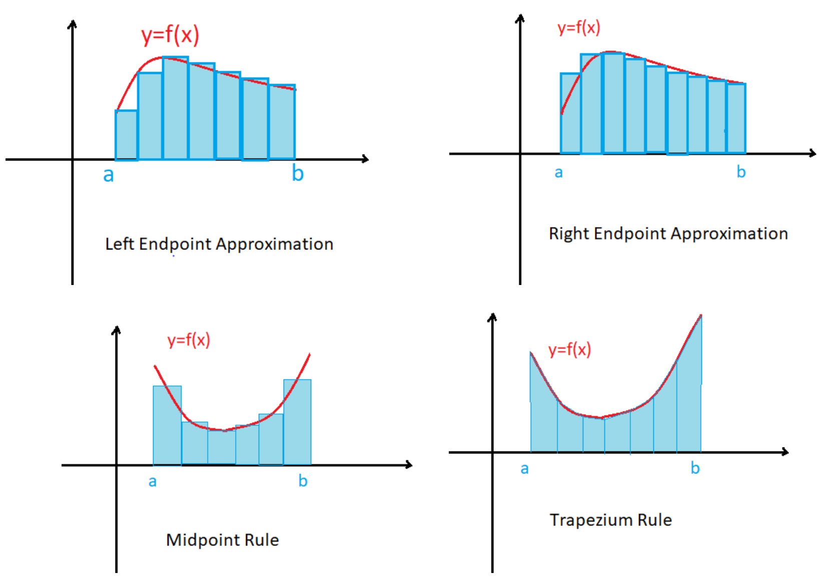
$$\int_a^b f(x)\,dx \approx (\sum_{i=1}^{n} f(x_i))\Delta x$$

Instead of using endpoints, we can also use midpoints of each interval for the approximation. Take $x_i^* = \bar{x}_i = \frac{x_{i-1}+x_i}{2}$, then the Riemann Sum is

$$M_n = \sum_{i=1}^{n} f(\bar{x}_i))\Delta x = (\sum_{i=1}^{n} f(\bar{x}_i))\frac{b-a}{n}$$

So we get **Midpoint Approximation**:

$$\int_a^b f(x)\,dx \approx (\sum_{i=1}^{n} f(\bar{x}_i))\Delta x$$



2 Trapezium Rule

Another commonly used approximation of integrals is the Trapezium Rule (also called the Trapezoid Rule). This method is not based on the idea of Riemann Sum, but to use a sequence of trapeziums to approximate the area on the xy-plane that represents the integral.

We first divide the interval $[a, b]$ evenly into n subintervals of equal length $\Delta x = \frac{b-a}{n}$, and let $x_i = a + i\Delta x$. We construct n trapeziums, with the i-th one based on the interval $[x_{i-1}, x_i]$, with the two vertical edges having top vertex on the graph of $y = f(x)$.

The area of the i-th trapezium is

$$\frac{f(x_{i-1}) + f(x_i)}{2}\Delta x$$

So the sum of these area is

$$\sum_{i=1}^{n} \frac{f(x_{i-1}) + f(x_i)}{2}\Delta x = (f(x_0)+2f(x_1)+2f(x_2)+...+2f(x_{i-2})+2f(x_{i-1})+f(x_i))\frac{b-a}{2n}$$

Then we use the sum of the area of these trapeziums to approximate the integral, which is called the **Trapezium Rule**:

$$\int_a^b f(x)\,dx \approx (f(x_0)+2f(x_1)+2f(x_2)+...+2f(x_{i-2})+2f(x_{i-1})+f(x_i))\frac{\Delta x}{2}$$

3 Simpson's Rule

A generalisation of the idea of Trapezium Rule is to use parabolas to interpolate the function instead of straight lines, and evaluate the area bounded by the parabolas.

Again, we divide the interval $[a, b]$ into n equal length subintervals with length $\Delta x = \frac{b-a}{n}$, but now we require n to be an even number. For an even number $0 \leq i \leq n-2$, we consider the parabola passing through the points $(x_i, f(x_i))$, $(x_{i+1}, f(x_{i+1}))$, $(x_{i+2}, f(x_{i+2}))$. To compute the area bounded by this parabola on the interval $[x_i, x_{i+2}]$, we can shift the parabola horizontally, and compute the area of the parabola $y = ax^2 + bx + c$ passing through the points $(-\Delta x, f(x_i))$, $(0, f(x_{i+1}))$, $(\Delta x, f(x_{i+2}))$ on the interval $[-\Delta x, \Delta x]$.

Plug in the three points to the equation of the parabola, we have

$$\begin{cases} f(x_i) = a(-\Delta x)^2 + b(-\Delta x) + c \\ f(x_{i+1}) = c \\ f(x_{i+2}) = a(\Delta x)^2 + b(\Delta x) + c \end{cases}$$

This implies

$$f(x_i) + 4f(x_{i+1}) + f(x_{i+2}) = 2a(\Delta x)^2 + 6c$$

The area is

$$\int_{-\Delta x}^{\Delta x} ax^2+bx+c\,dx = \frac{a}{3}x^3+\frac{b}{2}x^2+cx\Big|_{-\Delta x}^{\Delta x} = \frac{\Delta x}{3}(2a(\Delta x)^2+6c) = \frac{\Delta x}{3}(f(x_i)+4f(x_{i+1})+f(x_{i+2}))$$

Next we sum up the area bounded by each of these parabolas:

$$\frac{\Delta x}{3}((f(x_0) + 4f(x_1) + f(x_2)) + (f(x_0) + 4f(x_1) + f(x_2)) + ... + (f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)))$$
$$= \frac{\Delta x}{3}(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + ... + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n))$$

And this area provides an approximation of the integral $\int_a^b f(x)\,dx$, called the **Simpson's Rule**:

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{3}(f(x_0)+4f(x_1)+2f(x_2)+4f(x_3)+2f(x_4)+...+2f(x_{n-2})+4f(x_{n-1})+f(x_n))$$

4 Example

Let $f(x) = \ln x$ on $[1, 2]$, we know

$$\int_1^2 \ln x\,dx = \ln 4 - 1 \approx 0.386294$$

We will approximate $\int_1^2 \ln x\,dx$ by the methods described above, by taking $n = 5$, so $\Delta x = 0.2$, $x_0 = 1, x_1 = 1.2, x_2 = 1.4, x_3 = 1.6, x_4 = 1.8, x_5 = 2$

Left Endpoint Approximation:

$$\int_1^2 \ln x\,dx \approx (f(1) + f(1.2) + f(1.4) + f(1.6) + f(1.8)) \times 0.2 \approx 0.315317$$

Right Endpoint Approximation:

$$\int_1^2 \ln x \, dx \approx (f(1.2) + f(1.4) + f(1.6) + f(1.8) + f(2)) \times 0.2 \approx 0.453947$$

Midpoint Approximation:

$$\int_1^2 \ln x \, dx \approx (f(1.1) + f(1.3) + f(1.5) + f(1.7) + f(1.9)) \times 0.2 \approx 0.387124$$

Trapezium Rule:

$$.\int_1^2 \ln x \, dx \approx (f(1) + 2f(1.2) + 2f(1.4) + 2f(1.6) + 2f(1.8) + f(2)) \times 0.1 \approx 0.384633$$

For the Simpson's Rule, we take $n = 10$, so $\Delta x = 0.1$.
Simpson's Rule: and

$$\begin{aligned}
&\int_1^2 \ln x \, dx \\
\approx &(f(1) + 4f(1.1) + 2f(1.2) + 4f(1.3) + 2f(1.4) + 4f(1.5) + 2f(1.6) \\
&+ 4f(1.7) + 2f(1.8) + 4f(1.9) + f(2)) \times \frac{0.1}{3} \\
\approx &0.386294
\end{aligned}$$

5 Error for Approximation

Definition 1. The error of an approximation is defined to be the different between the accurate value and the approximated value. If we use A_n to denote an approximation with number of subintervals n, we denote the error by $E_A = \int_a^b f(x) \, dx - A_n$

The following Theorem follows from some theory in Numerical Analysis.

Theorem 2. *(i). f is a function on $[a, b]$ with $|f''(x)| \leq K$. Then the error for the Trapezium Rule is $|E_T| \leq \frac{K(b-a)^3}{12n^2}$, and the error for the Midpoint Rule is $|E_M| \leq \frac{K(b-a)^3}{24n^2}$.*

(ii). f is a function on $[a, b]$ with $|f^{(4)}(x)| \leq K$. Then the error for the Simpson's Rule is $|E_S| \leq \frac{K(b-a)^5}{180n^4}$

Example 3. *How large should we take n in order to make sure the error using Simpson's Rule for $\int_1^2 \ln x\, dx$ is within $0.00001 = 10^{-5}$?*

$f(x) = \ln x$, so $f^{(4)}(x) = -\frac{6}{x^4}$, so $|f^{(4)}| = |-\frac{6}{x^4}| \leq 6$.
We thus need $\frac{K(b-a)^5}{180n^4} = E_S \leq 10^{-6}$, taking $K = 6, a = 1, b = 2$:

$$\frac{6(2-1)^5}{180n^4} = \frac{1}{30n^4} \leq \frac{1}{10^5}$$

we get $n \geq \sqrt[4]{\frac{10^5}{30}} \approx 4.27$, so we need to take at least $n = 5$ to make sure the error is within 10^{-5}.