Approximating the Value of a Definite Integral

There are many ways to approximate the value of the integral $\int_a^b f(x)\,dx$ with a and b finite. We will use a simple approach. Subdivide the interval $[a, b]$ into a large number of subintervals of the same length. On each subinterval, approximate the integrand $f(x)$ by a function $s(x)$ which is so simple that we can explicitly evaluate its integral over the subinterval. The value of this integral is used as the approximation to the original integral over the subinterval. An approximation to the value of the integral over the entire interval $[a, b]$ is obtained by summing the approximations over all the subintervals.

To implement this idea, we first select a set of points $\{x_0, x_1, ..., x_N\}$ in the interval $[a, b]$. We require $x_0 = a$, $x_N = b$ and the points x_n to be equally spaced. Set $\Delta x = x_{n+1} - x_n$ which is the same positive constant for each subinterval. The set of points $\{x_0, x_1, ..., x_N\}$ is called a uniform partition of the interval $[a, b]$ with step size Δx. Next, we approximate the value of $\int f(x)\,dx$ over the subinterval $[x_n, x_{n+1}]$ by a number I_n. The number I_n is computed by integrating the approximating function $s(x)$ over the subinterval. The choice of $s(x)$ on the subinterval $[x_n, x_{n+1}]$ will be described below. The approximation to $\int_a^b f(x)\,dx$ is obtained by summing all the numbers I_n, $0 \le n \le N-1$.

Before we can actually compute an approximate value for an integral, we must explain how the approximating function $s(x)$ is to be chosen. The choice of $s(x)$ is dictated by the accuracy we want and how much work we are willing to do. To keep matters simple, we will assume that all computer calculations are exact and that there will always be enough computer memory available. For now, we will choose the simple, approximating functions $s(x)$ to be constant functions. The constant value used to approximate the integrand $f(x)$ over the subinterval $[x_n, x_{n+1}]$ will be either $f(x_n)$ or $f(x_{n+1})$. Setting $s(x) = f(x_n)$ on the subinterval $[x_n, x_{n+1}]$, we see

$$\int_{x_n}^{x_{n+1}} f(x)\,dx \approx \int_{x_n}^{x_{n+1}} f(x_n)\,dx = f(x_n)(x_{n+1} - x_n) = f(x_n)\Delta x_n$$

so $I_n = f(x_n)\Delta x_n$.

Now,

$$\int_a^b f(x)\,dx = \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} f(x)\,dx \approx \sum_{n=0}^{N-1} f(x_n)\Delta x$$

which is called the left sum approximation to $\int_a^b f(x)\,dx$ and is denoted LS. Setting $s(x) = f(x_{n+1})$ on the subinterval $[x_n, x_{n+1}]$ gives

$$\int_{x_n}^{x_{n+1}} f(x)\,dx \approx \int_{x_n}^{x_{n+1}} f(x_{n+1})\,dx = f(x_{n+1}) = I_n$$

now. Thus,

$$\int_a^b f(x)\,dx = \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} f(x)\,dx \approx \sum_{n=0}^{N-1} f(x_{n+1})\Delta x$$

which is called the right sum approximation to $\int_a^b f(x)\,dx$ and is denoted RS. Changing the index of summation shows that the right sum approximation can also be written as

$$\sum_{n=1}^{N} f(x_n)\Delta x$$

We always evaluate $f(x)$ at the left end of each subinterval or at the right end of each subinterval; we never mix the two.

Having defined the left and right sum approximations, we now have to estimate the error committed in using these approximations. Such estimates are important because they tell us how many subdivisions we must use to ensure a specified accuracy.

Monotone Functions

For concreteness, we assume $f(x)$ is nondecreasing on the interval $[a, b]$. Since $f(x)$ must be nondecreasing on each subinterval $[x_n, x_{n+1}]$, we have

$$f(x_n) \leq f(x) \leq f(x_{n+1})$$

so

$$f(x_n)\Delta x \leq \int_{x_n}^{x_{n+1}} f(x)\, dx \leq f(x_{n+1})\Delta x$$

Summing over n, we get

$$LS = \sum_{n=0}^{N-1} f(x_n)\Delta x \leq \int_a^b f(x)\, dx \leq \sum_{n=0}^{N-1} f(x_{n+1})\Delta x = RS$$

Thus, the left sum is a lower bound on the actual value of the integral and the right sum is an upper bound. To derive an error estimate, we look at the difference between LS and RS. Since all inequalities are reversed for nonincreasing functions, we use absolute values.

$$\begin{aligned} 0 \leq |LS - RS| &= \left| \sum_{n=0}^{N-1} f(x_n)\Delta x - \sum_{n=0}^{N-1} f(x_{n+1})\Delta x \right| \\ &= |(f(a) - f(b))\Delta x| = \frac{|f(b) - f(a)| \cdot |b - a|}{N} \end{aligned}$$

Since the actual value of the integral lies between LS and RS, the error committed in using either sum as an approximation to $\int_a^b f(x)\, dx$ cannot exceed $|LS - RS|$. Thus,

$$\left| \int_a^b f(x)\, dx - \{LS, RS\} \right| \leq \frac{|f(b) - f(a)| \cdot |b - a|}{N}$$

which is an error estimate for left sum and right sum approximations to $\int_a^b f(x)\, dx$ when the function $f(x)$ is monotone on $[a, b]$. To ensure that the error does not exceed a specified level tol, we require

$$\frac{|f(b) - f(a)| \cdot |b - a|}{N} \leq tol$$

or

$$N \geq \frac{|f(b) - f(a)| \cdot |b - a|}{tol} \qquad\qquad (f(x) \text{ monotone})$$

If we use at least this many subdivisions, we can be sure that the left and right sum approximations to $\int_a^b f(x)\,dx$ are correct to within an error that does not exceed tol.

Lipschitz Functions

To get started, we derive an error estimate for each subdivision $[x_n, x_{n+1}]$. An error estimate for the entire interval $[a, b]$ is obtained by adding all the error estimates for the subdivisions. The error estimate will be derived for the left sum; that for the right sums follows in exactly the same way. We begin by noting that

$$|f(x) - f(x_n)| \leq L|x - x_n| \leq L|x_{n+1} - x_n| = L\Delta x$$

for all x in $[x_n, x_{n+1}]$. Now,

$$\begin{aligned}
&\left| \int_{x_n}^{x_{n+1}} f(x)\,dx - \int_{x_n}^{x_{n+1}} f(x_n)\,dx \right| \\
&= \left| \int_{x_n}^{x_{n+1}} f(x) - f(x_n)\,dx \right| \\
&\leq \int_{x_n}^{x_{n+1}} |f(x) - f(x_n)|\,dx \\
&\leq L\Delta x \int_{x_n}^{x_{n+1}} dx \\
&= L(\Delta x)^2.
\end{aligned}$$

Since there are N subintervals in $[a, b]$, the total error committed in using the left sum approximation is bounded by

$$NL(\Delta x)^2 = \frac{L|b - a|^2}{N}$$

Then,

$$\left|\int_a^b f(x)\,dx - LS\right| \leq \frac{L\cdot|b-a|^2}{N}$$

is an error estimate for the left sum approximation for a Lipschitz function. The same estimate holds for a right sum in the same way. To ensure that the error does not exceed a specified level *tol*, we require

$$\frac{L\cdot|b-a|^2}{N} \leq tol$$

or

$$N \geq \frac{L\cdot|b-a|^2}{tol} \qquad\qquad (f(x)\ \ \text{Lipschitz})$$

If we use at least this many subdivisions, we can be sure that the left and right sum approximations are correct to within an error which does not exceed *tol*.

Continuously Differentiable Functions

As before, we first derive an error estimate for each subinterval and then sum them to get the final error estimate. The derivation of the error estimate in this case is based on a clever application of integration by parts. The estimate for the left sum will be derived. The change needed to derive the error estimate for the right sum will be mentioned. To start, let $g(x) = x - x_{n+1}$ on $[x_n, x_{n+1}]$. Note that $g'(x) \equiv 1$, $g(x_n) = -\Delta x$, and $g(x_{n+1}) = 0$. Thus,

$$\begin{aligned}
&\int_{x_n}^{x_{n+1}} f(x)\,dx \\
&= \int_{x_n}^{x_{n+1}} f(x)\cdot 1\,dx \\
&= \int_{x_n}^{x_{n+1}} f(x)g'(x)\,dx \\
&= f(x)g(x)\Big|_{x_n}^{x_{n+1}} - \int_{x_n}^{x_{n+1}} f'(x)g(x)\,dx
\end{aligned}$$

$$= 0 - f(x_n)(-\Delta x) - \int_{x_n}^{x_{n+1}} f'(x)(x - x_{n+1})\, dx$$

So,

$$\int_{x_n}^{x_{n+1}} f(x)\, dx = f(x_n)\Delta x - \int_{x_n}^{x_{n+1}} f'(x)(x - x_{n+1})\, dx$$

which is an exact equation. The first term on the right hand side is the left sum approximation to the integral on the left side of the equation. Ignoring the integral on the right hand side gets us back to the left sum approximation. We estimate the error, which is the integral we are ignoring, as follows.

$$\begin{aligned} |error| &= \left| - \int_{x_n}^{x_{n+1}} f'(x)(x - x_{n+1})\, dx \right| \\ &\leq \int_{x_n}^{x_{n+1}} |f'(x)| \cdot |x - x_n|\, dx \\ &\leq (\max |f'(x)|) \cdot (\Delta x)^2 \\ &= \|f'\|(\Delta x)^2 \end{aligned}$$

The notation $\|f'\|$ stands for $\max |f'(x)|$. By taking the max over the entire interval $[a, b]$ rather than the subinterval $[x_n, x_{n+1}]$, we get an estimate for the error which doesn't depend on the particular subinterval. Adding the error estimates for each of the N subintervals leads to the error bound $N\|f'\|(\Delta x)^2 = \dfrac{\|f'\| \cdot |b - a|^2}{N}$.

So,

$$\left| \int_a^b f(x)\, dx - LS \right| \leq \frac{\|f'\| \cdot |b - a|^2}{N}$$

is an error estimate for the left sum when $f(x)$ is continuously differentiable. Choosing $g(x) = x - x_n$ on $[x_n, x_{n+1}]$ leads to the same error estimate for the right sum approximation. To ensure that the error does not exceed a specified level tol, we require

$$\frac{\|f'\| \cdot |b - a|^2}{N} \leq tol$$

or

$$N \geq \frac{\|f'\| \cdot |b-a|^2}{tol} \qquad\qquad (f(x) \text{ continuously differentiable})$$

If we use at least this many subdivisions, we can be sure that the left and right sum approximations are correct to within an error that does not exceed tol. Note the similarity of this formula for N to that for a Lipschitz function.