Proof of Simpson's Formula

Simpson's method is

Given

$$I = \int_a^b f(x)\,dx$$

That is, given that we seek to integrate the function $f(x)$ between the limits a and b, start by finding values of h and n such that n is even and $b = a + nh$

Approximate I by

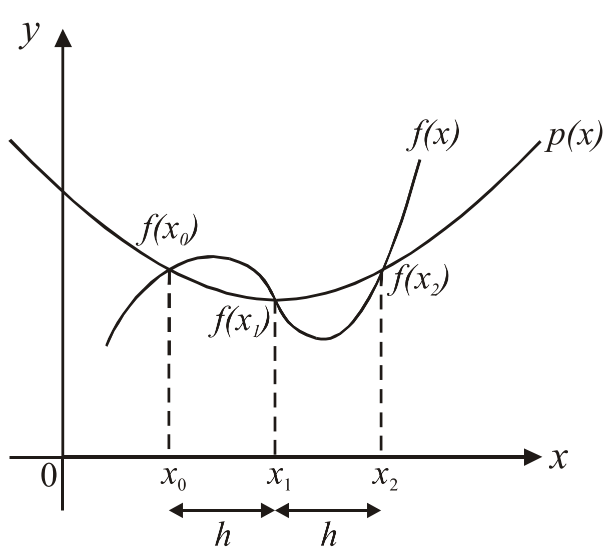
$$I \approx \tfrac{h}{3}\{f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 4f(x_{n-1}) + f(x_n)$$

where $x_i = a + ih$

In this unit we prove that this formula is correct. Firstly, we begin with the "simple" version of the formula – that is, when we approximate the area by using a single interval. Simpson's method uses three points for each interval and creates a quadratic function that runs through the three points.

To show that the approximation with three points is

$$I \approx \tfrac{h}{3}\left(f(x_0) + 4f(x_1) + f(x_2)\right)$$



Let the quadratic interpolating polynomial be

$$p(x) = a + bx + cx^2 \qquad (1)$$

We have

$$p(x_0) = f(x_0)$$
$$p(x_1) = f(x_1)$$
$$p(x_2) = f(x_2)$$

Substitution into (1) gives

$$f(x_0) = a + bx_0 + c(x_0)^2$$
$$f(x_1) = a + bx_1 + c(x_1)^2$$
$$f(x_2) = a + bx_2 + c(x_2)^2$$

Now

$$x_0 = x_1 - h$$
$$x_2 = x_1 + h$$

So this gives

$$f(x_0) = a + b(x_1 - h) + c(x_1 - h)^2$$
$$f(x_1) = a + b(x_1) + c(x_1)^2$$
$$f(x_2) = a + b(x_1 + h) + c(x_1 + h)^2$$

This is a bit unwieldy. Fortunately, there is a "get around" that simplifies things a little. This involves changing the variable. Suppose we define a variable

$$t = x - x_1$$

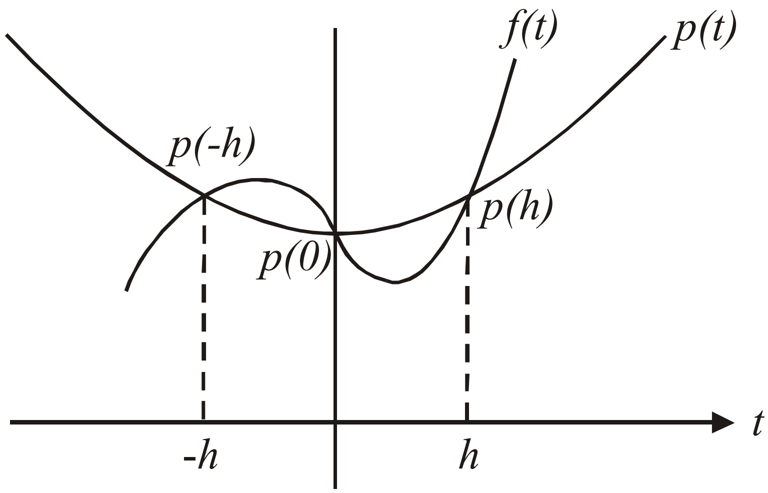
This variable takes the value 0 when $x = x_1$, the value $-h$ when $x = x_0$ and $+h$ when $x = x_2$.

The integral

$$\int_{-h}^{h} f(t)\,dt$$

is the same as

$$\int_{x_0}^{x_2} f(x)\,dx$$



With this change of variable the equations become

$$\begin{aligned} f(x_0) &= a + b(-h) + c(-h)^2 \\ f(x_1) &= a + \\ f(x_2) &= a + bh + ch^2 \end{aligned}$$

We solve this system of simultaneous equations by Gaussian row reduction.

$$\left[\begin{array}{ccc|c} a & -bh & ch^2 & f(x_0) \\ a & 0 & 0 & f(x_1) \\ a & bh & ch^2 & f(x_2) \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 0 & -bh & ch^2 & f(x_0) - f(x_1) \\ a & 0 & 0 & f(x_1) \\ 0 & bh & ch^2 & f(x_2) - f(x_1) \end{array}\right] \begin{array}{l} (1)-(2) \\ \\ (3)-(2) \end{array}$$

$$
\left[\begin{array}{ccc|c}
0 & 0 & 2ch^2 & \{f(x_0)-f(x_1)\}+\{f(x_2)-f(x_1)\} \\
a & 0 & 0 & f(x_1) \\
0 & 2bh & 0 & \{f(x_2)-f(x_1)\}+\{f(x_0)-f(x_1)\}
\end{array}\right]
\begin{array}{l}
(1)+(3) \\
\\
(3)-(1)
\end{array}
$$

Therefore,

$$
\begin{aligned}
a &= f(x_1) \\
b &= \tfrac{1}{2h^2}\left(f(x_2)-f(x_0)\right) \\
c &= \tfrac{1}{2h^2}\left\{f(x_0)+f(x_2)-2f(x_1)\right\}
\end{aligned}
$$

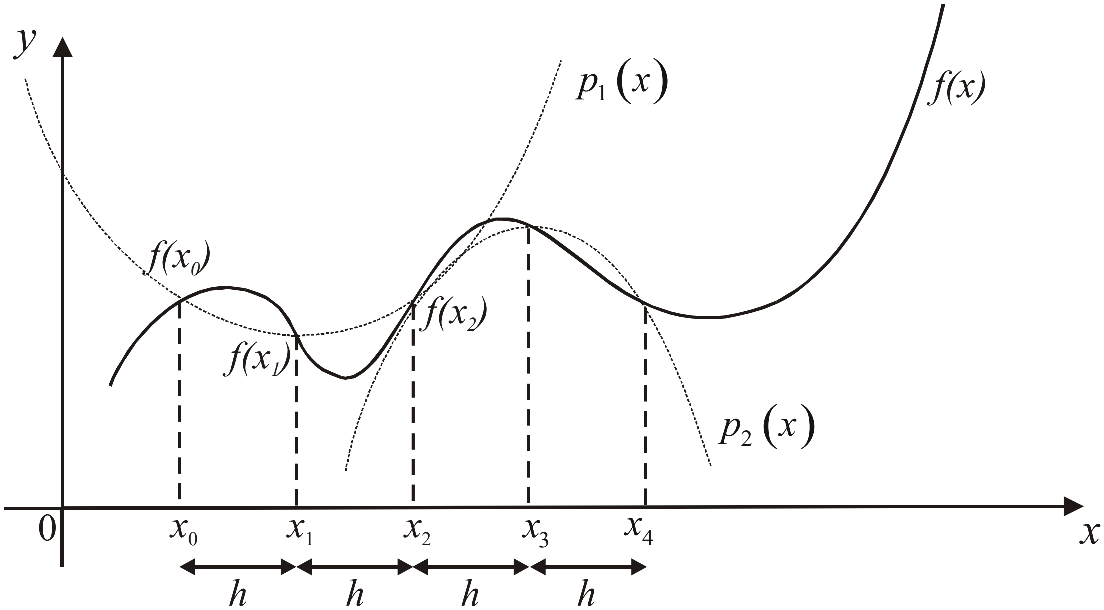
Hence the quadratic interpolating polynomial is

$$
q(t)=f(x_1)+\tfrac{1}{2h}\left(f(x_2)-f(x_0)\right)t+\tfrac{1}{2\mathrm{h}^2}\{f(x_0)+f(x_2)-2f(x_1)\}t^2
$$

Therefore,

$$
\begin{aligned}
f(x)dx &\approx \int_{x_0}^{x_2} p(x)dx \\
&= \int_{-h}^{h} q(t)dt \\
&= \int_{-h}^{h} f(x_1)+\tfrac{1}{2h}\left(f(x_2)-f(x_0)\right)t \ +\tfrac{1}{2h^2}\ \{f(x_0)+f(x_2)-2f(x_1)\}t^2 dt \\
&= \left[f(x_1)h+\tfrac{1}{2h}\left(f(x_2)-f(x_0)\right)\tfrac{t^2}{2}+\tfrac{1}{2h^2}\left(f(x_2)-2f(x_1)+f(x_0)\right)\tfrac{s^3}{3}\right] \\
&= \left(f(x_1)h+\tfrac{1}{2h}\left(f(x_2)-f(x_0)\right)\tfrac{h^2}{2}+\tfrac{1}{2h^2}\left(f(x_2)-2f(x_1)+f(x_0)\right)\tfrac{h^3}{3}\right)- \\
&\qquad \left(-f(x_1)h+\tfrac{1}{2h}\left(f(x_2)-f(x_0)\right)\tfrac{h^2}{2}+\tfrac{1}{2h^2}\left(f(x_2)-2f(x_1)+f(x_0)\right)\tfrac{h^3}{3}\right) \\
&= 2hf(x_1)+\tfrac{h}{3}\left(f(x_2)-2f(x_1)+f(x_0)\right) \\
&= \tfrac{h}{3}\left(6hf(x_1)+f(x_2)-2f(x_1)+f(x_0)\right) \\
&= \tfrac{h}{3}\left(f(x_2)+4f(x_1)+f(x_0)\right)
\end{aligned}
$$

This demonstrates the validity of the formula for three data points separated by two equal intervals. We will wish to use the formula in composite form - that is, with more than three data points. To do this we add together approximations of separate portions of the interval.



To illustrate this let us consider an approximation to the area under a curve $y = f(x)$, which is created by constructing two quadratic curves, one covering the first two intervals, the other covering the third and fourth intervals. Then the quadratic approximation to the first two intervals is

$$p_1(x) = \tfrac{h}{3}\left(f(x_2) + 4f(x_1) + f(x_0)\right)$$

And to the third and fourth intervals

$$p_2(x) = \tfrac{h}{3}\left(f(x_4) + 4f(x_3) + f(x_2)\right)$$

To the whole interval

$$\begin{aligned} p(x) &= p_1(x) + p_2(x) \\ &= \tfrac{h}{3}\left(f(x_2) + 4f(x_1) + 2f(x_2) + \tfrac{h}{3}\left(f(x_2) + 4f(x_3) - f(x_4)\right)\right) \\ &= \tfrac{h}{3}\left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + f(x_4)\right) \end{aligned}$$

Iterating this process for more and more segments gives the following description of Simpson's method in composite form

Given

$$I = \int_a^b f(x)\,dx$$

That is, given that we seek to integrate the function $f(x)$ between the limits *a* and *b*, start by finding values of *h* and *n* such that *n* is even and $b = a + nh$

Approximate *I* by

$$I \approx \tfrac{h}{3}\{f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 4f(x_{n-1}) + f(x_n)$$

where $x_i = a + ih$