Simpson's Rule Integration with MS Excel and Irregularly-spaced Data

Kenneth V. Cartwright, Ph.D. †

Abstract

A recent publication presented a method to numerically integrate irregularly-spaced data using Simpson's Rule. Unfortunately, this method is unsuitable for implementation in spreadsheets. To overcome this limitation, three alternative methods are suggested. Examples using MS Excel are given.

Keywords: Simpson's Rule, numerical integration using MS Excel, divided differences, irregularly-spaced data

Introduction

Quite often, it is necessary to integrate data (e.g. experimental) which are irregularly spaced. A recent publication [1] has shown how this can be accomplished using a computer language such as FORTRAN and Simpson's Rule. Unfortunately, as stated in [1], "Given the matrix inversions involved, this scheme not easily programmed with a spreadsheet." However, as is pointed out in this paper, the matrix conversions are not necessary and indeed integration using Simpson's Rule can be accomplished with a spreadsheet such as MS Excel for data points that are irregularly spaced in the abscissa coordinate. The authors of [2] have previously considered the equally-spaced case.

Statement of the problem

Given the N data points $\{(x_1, f(x_1)), (x_2, f(x_2)), ...(x_N, f(x_N))\}$, where N is an odd integer, find an estimate to the integral $I = \int_{x_1}^{x_N} f(x)dx$ using Simpson's rule. Assume that the abscissa values $\{x_1, x_2, ...x_N\}$ are not equally spaced. Note that in some cases, $f(x)$ for general x might be unknown, i.e., only $f(x_i)$, $i = 1, 2, ...N$, are known.

Previous solutions to the problem

Note that $I=\int_{x_1}^{x_N} f(x)dx=\int_{x_1}^{x_3} f(x)dx+\int_{x_3}^{x_5} f(x)dx...+\int_{x_{N-2}}^{x_N} f(x)dx.$ Hence, the total area is the sum of the individual areas computed using only three points, i.e. $I=I_3+I_5+...I_N$. Hence, in this section, without loss of generality, it is assumed that there are only three data points. For all methods of solution, Simpson's Rule requires that a parabola $f(x)=Ax^2+Bx+C$ is fitted to the points $\{(x_1,f(x_1)),(x_2,f(x_2)),(x_3,f(x_3))\}$. Hence, the following must be true:

$$Ax_1^2+Bx_1+C=f(x_1) \quad (1)$$

$$Ax_2^2+Bx_2+C=f(x_2) \quad (2)$$

$$Ax_3^2+Bx_3+C=f(x_3). \quad (3)$$

The area under the three-point parabolic segment is then given by

$$I=\int_{x_1}^{x_3} f(x)dx \approx \int_{x_1}^{x_3}\left(Ax^2+Bx+C\right)dx=\frac{A}{3}\left(x_3^3-x_1^3\right)+\frac{B}{2}\left(x_3^2-x_1^2\right)+C\left(x_3-x_1\right). \quad (4)$$

Solution Method I

In [1], it is assumed that Eq. (1), Eq. (2) and Eq. (3) can be written as a single matrix equation, given as

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{bmatrix}\begin{bmatrix} A \\ B \\ C \end{bmatrix}=\begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix}. \quad (5)$$

Hence, the parabola coefficients can be found from

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix}=\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{bmatrix}^{-1}\begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix}. \quad (6)$$

Unfortunately, the matrix inversion makes it difficult to implement Eq. (6) in spreadsheets, as noted earlier.

Solution Method II

The author of [3] assumed that the three points given are $\{(-h, f(-h)), (0, f(0)), (k, f(k))\}$. Using divided differences, the said author proved that

$$I = \int_{-h}^{k} f(x)dx \approx \frac{1}{6}(h+k)\left[\left\{2-\frac{k}{h}\right\} f(-h) + \frac{(h+k)^2}{hk} f(0) + \left\{2-\frac{h}{k}\right\} f(k)\right]. \tag{7}$$

However, we wish to compute $\int_{x_1}^{x_3} f(x)dx$. This case was not addressed in [3].

Nonetheless, we can use Eq. (7) if we replace $f(-h)$ by $f(x_1)$, $f(0)$ by $f(x_2)$, $f(k)$ by $f(x_3)$, k by $x_3 - x_2$, and h by $x_2 - x_1$. Hence,

$$I = \int_{x_1}^{x_3} f(x)dx \approx \frac{1}{6}(x_3 - x_1)\left[\begin{array}{l}\left\{2-\frac{x_3-x_2}{x_2-x_1}\right\} f(x_1) + \frac{(x_3-x_1)^2}{(x_3-x_2)(x_2-x_1)} f(x_2) \\ +\left\{2-\frac{x_2-x_1}{x_3-x_2}\right\} f(x_3)\end{array}\right]. \tag{8}$$

Note that if the distances between points are equal, i.e., $d = x_3 - x_2 = x_2 - x_1$, Eq. (8) becomes $I \approx \frac{d}{3}\left[f(x_1) + 4f(x_2) + f(x_3)\right]$, i.e., the well-known Simpson's Rule formula.

Fortunately, we can easily use Eq. (8) in a spreadsheet.

Additional solutions to the problem

In this section, we present two additional methods of determining the parabola coefficients: one of these methods use Lagrange Interpolation and the other solves the simultaneous equations given by Eq. (1), Eq. (2) and Eq. (3) without explicit matrix inversion.

Solution Method III

It is known that Lagrange Interpolation can be used to derive Simpson's Rule for the equally-spaced case. In this sub-section, it will be seen that it can also be used for the unequally-spaced case as well.

By the method of Lagrange Interpolation,

$$
\begin{aligned}
f(x) &= \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} f(x_1) + \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} f(x_2) + \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} f(x_3) \\
&= \frac{x^2-(x_2+x_3)x+x_2x_3}{(x_1-x_2)(x_1-x_3)} f(x_1) + \frac{x^2-(x_1+x_3)x+x_1x_3}{(x_2-x_1)(x_2-x_3)} f(x_2) \\
&\quad + \frac{x^2-(x_1+x_2)x+x_1x_2}{(x_3-x_1)(x_3-x_2)} f(x_3).
\end{aligned}
\tag{9}
$$

Comparing Eq. (9) with $f(x) = Ax^2 + Bx + C$ gives

$$
\begin{aligned}
A &= \frac{1}{(x_1-x_2)(x_1-x_3)} f(x_1) + \frac{1}{(x_2-x_1)(x_2-x_3)} f(x_2) + \frac{1}{(x_3-x_1)(x_3-x_2)} f(x_3) \\
&= \frac{1}{(x_2-x_1)(x_3-x_1)} f(x_1) - \frac{1}{(x_2-x_1)(x_3-x_2)} f(x_2) + \frac{1}{(x_3-x_1)(x_3-x_2)} f(x_3),
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
B &= -\frac{(x_2+x_3)}{(x_1-x_2)(x_1-x_3)} f(x_1) - \frac{(x_1+x_3)}{(x_2-x_1)(x_2-x_3)} f(x_2) - \frac{(x_1+x_2)}{(x_3-x_1)(x_3-x_2)} f(x_3) \\
&= -\frac{(x_2+x_3)}{(x_2-x_1)(x_3-x_1)} f(x_1) + \frac{(x_1+x_3)}{(x_2-x_1)(x_3-x_2)} f(x_2) - \frac{(x_1+x_2)}{(x_3-x_1)(x_3-x_2)} f(x_3),
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
C &= \frac{x_2x_3}{(x_1-x_2)(x_1-x_3)} f(x_1) + \frac{x_1x_3}{(x_2-x_1)(x_2-x_3)} f(x_2) + \frac{x_1x_2}{(x_3-x_1)(x_3-x_2)} f(x_3) \\
&= \frac{x_2x_3}{(x_2-x_1)(x_3-x_1)} f(x_1) - \frac{x_1x_3}{(x_2-x_1)(x_3-x_2)} f(x_2) + \frac{x_1x_2}{(x_3-x_1)(x_3-x_2)} f(x_3).
\end{aligned}
\tag{12}
$$

Using Eq. (10), Eq. (11), Eq. (12) and Eq. (4), the integral is easily evaluated in a spreadsheet.

Solution Method IV

In this final method, we solve Eq. (1), Eq. (2) and Eq. (3) algebraically.

Subtracting Eq. (1) from Eq. (2) gives

$$A\left(x_2^2 - x_1^2\right) + B\left(x_2 - x_1\right) = f\left(x_2\right) - f\left(x_1\right). \tag{13}$$

Solving Eq. (13) gives

$$B = \frac{f\left(x_2\right) - f\left(x_1\right)}{\left(x_2 - x_1\right)} - A\left(x_1 + x_2\right). \tag{14}$$

Additionally, subtracting Eq. (2) from Eq. (3) gives

$$A\left(x_3^2 - x_2^2\right) + B\left(x_3 - x_2\right) = f\left(x_3\right) - f\left(x_2\right). \tag{15}$$

Solving Eq. (15) produces

$$B = \frac{f\left(x_3\right) - f\left(x_2\right)}{\left(x_3 - x_2\right)} - A\left(x_2 + x_3\right). \tag{16}$$

Setting Eq. (14) and Eq. (16) equal and rearranging gives

$$\begin{aligned} A &= \frac{\dfrac{f\left(x_3\right) - f\left(x_2\right)}{\left(x_3 - x_2\right)} - \dfrac{f\left(x_2\right) - f\left(x_1\right)}{\left(x_2 - x_1\right)}}{\left(x_3 - x_1\right)} \\ &= \frac{f\left(x_3\right) - f\left(x_2\right)}{\left(x_3 - x_2\right)\left(x_3 - x_1\right)} - \frac{f\left(x_2\right) - f\left(x_1\right)}{\left(x_2 - x_1\right)\left(x_3 - x_1\right)}. \end{aligned} \tag{17}$$

Substituting Eq. (17) into Eq. (14) or Eq. (16) gives the desired formula for B.

Finally, rearranging Eq. (1) produces

$$C = f\left(x_1\right) - Ax_1^2 - Bx_1. \tag{18}$$

Using Eq. (17), Eq. (14) or Eq. (16), Eq. (18) and then Eq. (4), the integral is easily computed in a spreadsheet.

Note that Eq. (17) must equal Eq. (10), Eq. (16) must equal Eq. (11), and Eq. (12) must equal Eq. (18). It is straightforward to show this algebraically; alternatively, the MS Excel computations in the tables of the next section show that this is the case.

Implementation in MS Excel

In this section, implementation in MS Excel examples will be given.

Example 1

Let the points be $\{(1,1),(1.25,1.5625),(1.75,3.0625)\}$. These points were taken from the function $f(x)=x^2$. Hence, the integral is exactly $1.75^3/3-1/3=1.453125$. Furthermore, integration with Simpson's Rule should give this exact answer, as well. Fortunately, it does as is shown in cell F3 of Table I, cell I3 of Table II and cell I3 of Table III.

Table I. Implementation of Example 1 using Eq. (8) in MS Excel.

	A	B	C	D	E	F
1	1	1				
2	1.25	1.5625				
3	1.75	3.0625	0.25	0.5	0.75	1.453125
4						
5	C3=A2-A1					
6	D3=A3-A2					
7	E3=A3-A1					
8	F3=E3/6*((2-D3/C3)*B1+E3*E3/D3/C3*B2+(2-C3/D3)*B3)					
9						
10	*x* values are in column A, *f(x)* in column B.					

Note that cell F3 of Table I is computing Eq. (8).

Table II. Implementation of Example 1 using Solution Method III in MS Excel.

	A	B	C	D	E	F	G	H	I
1	1	1							
2	1.25	1.5625							
3	1.75	3.0625	0.25	0.5	0.75	1	0	0	1.453125
4									
5	C3=A2-A1								
6	D3=A3-A2								
7	E3=A3-A1								
8	F3=B1/C3/E3-B2/C3/D3+B3/D3/E3								
9	G3=-B1/C3/E3*(A2+A3)+B2/C3/D3*(A1+A3)-B3/D3/E3*(A1+A2)								
10	H3=B1/C3/E3*(A2*A3)-B2/C3/D3*(A1*A3)+B3/D3/E3*(A1*A2)								
11	I3=F3/3*(A3*A3*A3-A1*A1*A1)+G3/2*(A3*A3-A1*A1)+H3*(A3-A1)								
12									
13	*x* values are in column A, *f(x)* in column B.								

In Table II, cell F3 is computing Eq. (10), cell G3 is computing Eq. (11), cell H3 is computing Eq. (12) and cell I3 computes Eq. (4).

Table III. Implementation of Example 1 using Solution Method IV in MS Excel.

	A	B	C	D	E	F	G	H	I
1	1	1							
2	1.25	1.5625							
3	1.75	3.0625	0.25	0.5	0.75	1	0	0	1.453125
4									
5	C3=A2-A1								
6	D3=A3-A2								
7	E3=A3-A1								
8	F3=((B3-B2)/D3-(B2-B1)/C3)/E3								
9	G3=(B3-B2)/D3-F3*(A2+A3)								
10	H3=B1-F3*A1*A1-G3*A1								
11	I3=F3/3*(A3*A3*A3-A1*A1*A1)+G3/2*(A3*A3-A1*A1)+H3*(A3-A1)								
12									
13	*x* values are in column A, *f(x)* in column B.								

In Table III, cell F3 is computing Eq. (17), cell G3 is computing Eq. (16), cell H3 is computing Eq. (18) and cell I3 computes Eq. (4).

Example 2

In this example, we will use nine points which are taken from $f(x) = \sin x$. Hence, the integral is $1 - \cos(0.9) = 0.3783900$; whereas, Simpson's Rule gives this as 0.3783929, as shown in Tables IV, V and VI below.

Table IV. Implementation of Example 2 using Eq. (8) in MS Excel.

	A	B	C	D	E	F	G
1	0	0					
2	0.1	0.0998334					
3	0.19	0.1888589	0.1	0.09	0.19	0.01799672	
4	0.33	0.324043					
5	0.4	0.3894183	0.14	0.07	0.21	0.06095187	
6	0.55	0.5226872					
7	0.69	0.6365372	0.15	0.14	0.29	0.14981824	
8	0.74	0.6742879					
9	0.9	0.7833269	0.05	0.16	0.21	0.14962611	0.3783929
10							
11	C3=A2-A1						
12	D3=A3-A2						
13	E3=A3-A1						
14	F3=E3/6*((2-D3/C3)*B1+E3*E3/D3/C3*B2+(2-C3/D3)*B3)						
15	G9=SUM(F3,F5,F7,F9)						
16							
17	*x* values are in column A, *f(x)* in column B.						

Note for Table IV,

(i) cells C3, D3, E3 and F3 are copied to cells CM, DM, EM and FM, where M=5,7 and 9; and

(ii) $F3 \approx \int_0^{0.19} f(x)dx,\ F5 \approx \int_{0.19}^{0.4} f(x)dx,\ F7 \approx \int_{0.4}^{0.69} f(x)dx$ and $F9 \approx \int_{0.69}^{0.9} f(x)dx.$

Hence, $G9 = F3 + F5 + F7 + F9 \approx \int_0^{0.9} f(x)dx.$

Table V. Implementation of Example 2 using Solution Method III in MS Excel.

	A	B	C	D	E	F	G	H	I
1	0	0							
2	0.1	0.0998334							
3	0.19	0.1888589	0.1	0.09	0.19	-0.048222	1.003156369	0	0.0179967
4	0.33	0.324043							
5	0.4	0.3894183	0.14	0.07	0.21	-0.1507995	1.044016698	-0.0040604	0.0609519
6	0.55	0.5226872							
7	0.69	0.6365372	0.15	0.14	0.29	-0.2594665	1.134952442	-0.023048	0.1498182
8	0.74	0.6742879							
9	0.9	0.7833269	0.05	0.16	0.21	-0.3500993	1.255656571	-0.0631836	0.1496261
10									
11									0.3783929
12	C3=A2-A1								
13	D3=A3-A2								
14	E3=A3-A1								
15	F3=B1/C3/E3-B2/C3/D3+B3/D3/E3								
16	G3=-B1/C3/E3*(A2+A3)+B2/C3/D3*(A1+A3)-B3/D3/E3*(A1+A2)								
17	H3=B1/C3/E3*(A2*A3)-B2/C3/D3*(A1*A3)+B3/D3/E3*(A1*A2)								
18	I3=F3/3*(A3*A3*A3-A1*A1*A1)+G3/2*(A3*A3-A1*A1)+H3*(A3-A1)								
19	I11=SUM(I3,I5,I7,I9)								
20									
21	*x* values are in column A, *f(x)* in column B.								

Note that in Tables V and VI,

(i) cells C3, D3, E3, F3, G3, H3 and I3 are copied to cells CM, DM, EM, FM, GM, HM and IM where M=5,7 and 9;

(ii) the entries for Tables V and VI have the same entries. This is because the equations for the parabola coefficients in Table V are equal to the corresponding equations in Table VI, as noted earlier; and

(iii) $I3 \approx \int_0^{0.19} f(x)dx,\ I5 \approx \int_{0.19}^{0.4} f(x)dx,\ I7 \approx \int_{0.4}^{0.69} f(x)dx$ and $I9 \approx \int_{0.69}^{0.9} f(x)dx.$

Hence, $I11 = I3 + I5 + I7 + I9 \approx \int_0^{0.9} f(x)dx.$

Table VI. Implementation of Example 2 using Solution Method IV in MS Excel.

	A	B	C	D	E	F	G	H	I
1	0	0							
2	0.1	0.0998334							
3	0.19	0.1888589	0.1	0.09	0.19	-0.048222	1.003156369	0	0.017996721
4	0.33	0.324043							
5	0.4	0.3894183	0.14	0.07	0.21	-0.1507995	1.044016698	-0.0040604	0.060951869
6	0.55	0.5226872							
7	0.69	0.6365372	0.15	0.14	0.29	-0.2594665	1.134952442	-0.023048	0.149818239
8	0.74	0.6742879							
9	0.9	0.7833269	0.05	0.16	0.21	-0.3500993	1.255656571	-0.0631836	0.149626108
10									
11									0.3783929
12	C3=A2-A1								
13	D3=A3-A2								
14	E3=A3-A1								
15	F3=((B3-B2)/D3-(B2-B1)/C3)/E3								
16	G3=(B3-B2)/D3-F3*(A2+A3)								
17	H3=B1-F3*A1*A1-G3*A1								
18	I3=F3/3*(A3*A3*A3-A1*A1*A1)+G3/2*(A3*A3-A1*A1)+H3*(A3-A1)								
19	I11=SUM(I3,I5,I7,I9)								
20									
21	*x* values are in column A, *f(x)* in column B.								

Conclusion

Three different methods have been presented that allow the numerical integration of irregularly-spaced data. These methods are suitable for implementation in spreadsheets. Examples using MS Excel were given.

† *Kenneth V. Cartwright, Ph.D.*, University of The Bahamas, Nassau, Bahamas.

References

[1] Reed, B. C. (2014). Numerically Integrating Irregularly-spaced (x, y) Data. *The Mathematics Enthusiast*, *11*(3), 643-648.

[2] El-Gebeily, M., & Yushau, B. (2007). Numerical Methods with MS Excel. *The Mathematics Enthusiast*, *4*(1), 84-92.

[3] Shklov, N. (1960). Simpson's Rule for Unequally Spaced Ordinates. *The American Mathematical Monthly, 67*(10), 1022-1023.