Simpson's Rule Cumulative 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 and a spreadsheet. The method did not explicitly show how cumulative integration using Simpson's Rule may be computed, as required in some applications. This paper removes that shortcoming by the presentation of two different methods to achieve this goal. Examples using MS Excel are given.

Keywords: Simpson's Rule, numerical integration, cumulative numerical integration, integration using spreadsheets, irregularly-spaced data

Introduction

Often, it is necessary to cumulatively integrate data (e.g. experimental) which are irregularly spaced. For example, Foss's method [1] (and its modification [2]) of fitting exponential functions to data or the method by Jacquelin [3] that fits many different functions to data, require cumulative numerical integration. The authors of [2] and [3] suggest the use of the trapezoidal rule to accomplish this. However, in this paper, it is shown how this cumulative numerical integration can be accomplished using Simpson's Rule with MS Excel and irregularly-spaced data. In fact, two different methods will be given. The case with regularly-spaced data using a FORTRAN computer program has been previously considered by Blake [4], whereas [5] did the same using a spreadsheet and a different algorithm. Also, it should be mentioned that Cartwright [6] has shown how Simpson's Rule ordinary numerical integration can be done with spreadsheets such as Excel with irregularly spaced data. However, the method of [6] did not explicitly show how cumulative integration can be accomplished as required in some applications.

Statement of the problem

Given the N data points $\{(x_1, f(x_1)),(x_2, f(x_2)),....(x_N, f(x_N))\}$, find estimates to the integrals $I_i = \int_{x_1}^{x_i} f(x)dx$, where $i = 1,2,3,...N$, 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 $f(x_i)$, $i = 1, 2, \ldots N$, are known.

Background

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:

$$Ax_1^2 + Bx_1 + C = f(x_1) \tag{1}$$

$$Ax_2^2 + Bx_2 + C = f(x_2) \tag{2}$$

$$Ax_3^2 + Bx_3 + C = f(x_3). \tag{3}$$

Solving these equations, as shown by Cartwright [6], gives

$$\begin{aligned} A &= \frac{f(x_3) - f(x_2)}{(x_3 - x_2)(x_3 - x_1)} - \frac{f(x_2) - f(x_1)}{(x_2 - x_1)(x_3 - x_1)}, \\ B &= \frac{f(x_3) - f(x_2)}{(x_3 - x_2)} - A(x_2 + x_3), \\ C &= f(x_1) - Ax_1^2 - Bx_1. \end{aligned} \tag{4}$$

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

$$I_3 = \int_{x_1}^{x_3} f(x)dx \approx \int_{x_1}^{x_3} \left(Ax^2 + Bx + C\right)dx = \frac{A}{3}\left(x_3^3 - x_1^3\right) + \frac{B}{2}\left(x_3^2 - x_1^2\right) + C(x_3 - x_1). \tag{5}$$

As shown by Cartwright [6], Eq. (5) can also be written as

$$I_3 \approx \frac{(x_3 - x_1)}{6}\left[\left\{2 - \frac{x_3 - x_2}{x_2 - x_1}\right\} f(x_1) + \frac{(x_3 - x_1)^2}{(x_3 - x_2)(x_2 - x_1)} f(x_2) + \left\{2 - \frac{x_2 - x_1}{x_3 - x_2}\right\} f(x_3)\right]. \tag{6}$$

Also,

$$I_2 = \int_{x_1}^{x_2} f(x)dx \approx \int_{x_1}^{x_2} \left(Ax^2 + Bx + C\right)dx = \frac{A}{3}\left(x_2^3 - x_1^3\right) + \frac{B}{2}\left(x_2^2 - x_1^2\right) + C(x_2 - x_1). \tag{7}$$

Furthermore, using Eq. (4) and symbolic math programming in MATLAB (please see the Appendix), Eq. (7) can be shown to be given by

$$I_2 \approx \frac{(x_2-x_1)}{6}\left[\begin{array}{l}\left\{3-\dfrac{x_2-x_1}{x_3-x_1}\right\} f\left(x_1\right)+\left\{3+\dfrac{\left(x_2-x_1\right)^2}{\left(x_3-x_2\right)\left(x_3-x_1\right)}+\dfrac{x_2-x_1}{x_3-x_1}\right\} f\left(x_2\right) \\ -\dfrac{\left(x_2-x_1\right)^2}{\left(x_3-x_2\right)\left(x_3-x_1\right)} f\left(x_3\right)\end{array}\right]. \tag{8}$$

Note that the correctness of Eq. (8) will be demonstrated in the examples below.

Previous solutions to the problem for equally-spaced data

If the distances between points are equal, i.e., $d = x_3 - x_2 = x_2 - x_1$, Eq. (5) and Eq. (6) become the well-known Simpson's Rule

$$I_3 = \int_{x_1}^{x_3} f(x)dx \approx \frac{d}{3}\left(f\left(x_1\right)+4f\left(x_2\right)+f\left(x_3\right)\right). \tag{9}$$

However, what about I_2? Blake [4] has shown that this is equal to

$$I_2 \approx \frac{d}{3}\left(\frac{5}{4}f\left(x_1\right)+2f\left(x_2\right)-\frac{1}{4}f\left(x_3\right)\right). \tag{10}$$

It is rewarding to note that Eq. (8) becomes Eq. (10) for equally-spaced data, as it should.

Furthermore, Blake [4] provides an algorithm that performs cumulative numerical integration with a modified Simpson's Rule and equally-spaced data. This algorithm was programmed with the FORTRAN computer language but can in fact be made to work in spreadsheets. However, another algorithm that is well-known to engineers [5], [7] will be used in this paper: indeed the authors of [5] give the details of this method in Eq. (6) of [5], which is stated in Eq. (11) below:

$$I_k = I_{k-2} + \frac{d}{3}\left(f\left(x_{k-2}\right)+4f\left(x_{k-1}\right)+f\left(x_k\right)\right) \text{ for } 3 \le k \le N, \tag{11}$$

where $I_1 = 0$ and I_2 is given by Eq. (10).

In the next section, this algorithm will be extended to allow irregularly-spaced data.

Extension of Eq. (11) for irregularly-spaced data

Note that Eq. (11) can be generalized to

$$I_k = I_{k-2} + S_k, \text{ for } 3 \le k \le N, \tag{12}$$

where $I_1 = 0$, I_2 is given by Eq. (7) or Eq. (8) and S_k is the area under the parabola $f(x) = Ax^2 + Bx + C$ that is fitted to the points $\{(x_{k-2}, f(x_{k-2})), (x_{k-1}, f(x_{k-1})), (x_k, f(x_k))\}$.

Hence, the following must be true:

$$Ax_{k-2}^2 + Bx_{k-2} + C = f(x_{k-2}) \tag{13}$$

$$Ax_{k-1}^2 + Bx_{k-1} + C = f(x_{k-1}) \tag{14}$$

$$Ax_k^2 + Bx_k + C = f(x_k). \tag{15}$$

Solving these equations gives

$$\begin{aligned} A &= \frac{f(x_k) - f(x_{k-1})}{(x_k - x_{k-1})(x_k - x_{k-2})} - \frac{f(x_{k-1}) - f(x_{k-2})}{(x_{k-1} - x_{k-2})(x_k - x_{k-2})}, \\ B &= \frac{f(x_k) - f(x_{k-1})}{(x_k - x_{k-1})} - A(x_{k-1} + x_k), \\ C &= f(x_{k-2}) - Ax_{k-2}^2 - Bx_{k-2}. \end{aligned} \tag{16}$$

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

$$\begin{aligned} S_k &= \int_{x_{k-2}}^{x_k} f(x)dx \\ &\approx \int_{x_{k-2}}^{x_k} \left(Ax^2 + Bx + C\right)dx = \frac{A}{3}\left(x_k^3 - x_{k-2}^3\right) + \frac{B}{2}\left(x_k^2 - x_{k-2}^2\right) + C(x_k - x_{k-2}). \end{aligned} \tag{17}$$

Alternatively, Eq. (6) can be generalized to get

$$S_k \approx \frac{(x_k - x_{k-2})}{6}\left[\begin{array}{l} \left\{2 - \dfrac{x_k - x_{k-1}}{x_{k-1} - x_{k-2}}\right\} f(x_{k-2}) + \dfrac{(x_k - x_{k-2})^2}{(x_k - x_{k-1})(x_{k-1} - x_{k-2})} f(x_{k-1}) \\ + \left\{2 - \dfrac{x_{k-1} - x_{k-2}}{x_k - x_{k-1}}\right\} f(x_k) \end{array}\right]. \tag{18}$$

Note that Eq. (17) and Eq. (18) are valid for $3 \le k \le N$.

For comparison purposes in the next section, cumulative numerical integration results using the trapezium rule will also be given: the relevant equation is (see, e.g. [4] or [7])

$$I_k = I_{k-1} + 0.5\left(x_k - x_{k-1}\right)\left(f\left(x_{k-1}\right) + f\left(x_k\right)\right), \text{ for } 2 \le k \le N, \tag{19}$$

and $I_1 = 0$.

Implementation in MS Excel

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

Example 1

For this first example, let the points be taken from $f(x) = x^2$. Hence, the theoretical cumulative integral is given by $I = \int_0^x x^2 dx = \frac{x^3}{3}$. Using Eq. (7), Eq. (12), Eq. (16) and Eq. (17) (which will be called Method I), Simpson's Rule cumulative numerical integration was performed and the results are seen in Column I of Table I. The theoretical results are given in Column J. Clearly, the theoretical and numerical results are the same as they must be for Simpson's Rule for $f(x) = x^2$. Furthermore, the cumulative numerical integration using the trapezium rule results are given in Column K. Clearly, these latter results are poorer, as expected.

Example 2

For this second example, let the points be taken from $f(x) = \sin x$. Hence, the theoretical cumulative integral is given by $I = \int_0^x \sin dx = 1 - \cos x$. Using Eq. (7), Eq. (12), Eq. (16) and Eq. (17) (Method I), Simpson's Rule cumulative numerical integration was performed and the results are seen in Column I of Table II. The theoretical results are given in Column J. Furthermore, the cumulative numerical integration using the trapezium rule results are given in Column K. Clearly, Simpson's Rule cumulative integration gives answers closer to the theoretical values.

Table I. Implementation of Example 1 using Eq. (7), Eq. (12), Eq. (16) and Eq. (17) (Method I) in MS Excel.

	A	B	C	D	E	F	G	H	I	J	K
1	0	0							0	0	0
2	0.1	0.01							0.000333	0.000333	0.0005
3	0.19	0.0361	0.1	0.09	0.19	1	0	0	0.002286	0.002286	0.002575
4	0.33	0.1089	0.09	0.14	0.23	1	0	-1.7E-18	0.011979	0.011979	0.012725
5	0.4	0.16	0.14	0.07	0.21	1	0	-3.5E-17	0.021333	0.021333	0.022136
6	0.55	0.3025	0.07	0.15	0.22	1	9.99E-16	-2E-16	0.055458	0.055458	0.056824
7	0.69	0.4761	0.15	0.14	0.29	1	0	-5.6E-17	0.109503	0.109503	0.111326
8	0.74	0.5476	0.14	0.05	0.19	1	0	0	0.135075	0.135075	0.136918
9	0.9	0.81	0.05	0.16	0.21	1	0	-2.2E-16	0.243	0.243	0.245526
10											
11											
12	C3=A2-A1								K2=K1+0.5*(A2-A1)*(B1+B2)		
13	D3=A3-A2									J1=A!*A1*A1/3	
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)+I1										
19	I2=F3/3*(A2*A2*A2-A1*A1*A1)+G3/2*(A2*A2-A1*A1)+H3*(A2-A1)										
20											
21	*x* values are in column A, *f(x)* in column B.										

Table II. Implementation of Example 2 using Eq. (7), Eq. (12), Eq. (16) and Eq. (17) (Method I) in MS Excel.

	A	B	C	D	E	F	G	H	I	J	K
1	0	0							0	0	0
2	0.1	0.099833							0.00499971	0.00499583	0.00499167
3	0.19	0.188859	0.1	0.09	0.19	-0.04822	1.003156	0	0.01799672	0.01799576	0.01798282
4	0.33	0.324043	0.09	0.14	0.23	-0.10248	1.018892	-0.00103	0.05395332	0.05395766	0.05388596
5	0.4	0.389418	0.14	0.07	0.21	-0.1508	1.044017	-0.00406	0.07894859	0.07893901	0.07885711
6	0.55	0.522687	0.07	0.15	0.22	-0.2067	1.084823	-0.01144	0.14746047	0.14747548	0.14726503
7	0.69	0.636537	0.15	0.14	0.29	-0.25947	1.134952	-0.02305	0.22876683	0.22875399	0.22841073
8	0.74	0.674288	0.14	0.05	0.19	-0.30631	1.193041	-0.04083	0.26152327	0.26153144	0.26118136
9	0.9	0.783327	0.05	0.16	0.21	-0.3501	1.255657	-0.06318	0.37839294	0.37839003	0.37779055
10											
11											
12	C3=A2-A1								K2=K1+0.5*(A2-A1)*(B1+B2)		
13	D3=A3-A2									J1=1-COS(A1)	
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)+I1										
19	I2=F3/3*(A2*A2*A2-A1*A1*A1)+G3/2*(A2*A2-A1*A1)+H3*(A2-A1)										
20											
21	*x* values are in column A, *f(x)* in column B.										

Note to create Table I and Table II:

(i) $\mathrm{C3} = x_2 - x_1, \mathrm{D3} = x_3 - x_2$ and $\mathrm{E3} = x_3 - x_1$.
(ii) $\mathrm{F3} = A$ from Eq. (4), $\mathrm{G3} = B$ from Eq. (4) and $\mathrm{H3} = C$ from Eq. (4).
(iii) I2 calculates Eq. (7).
(iv) I3 calculates Eq. (5) plus I1.
(v) Cells C3, D3, E3, F3, G3, H3 and I3 are copied and pasted to partially

fill in rows 4 through 9.

(vi) K2 calculates Eq. (19).

(vii) K2 is copied and pasted into cells K3:K9.

Example 3

For this third example, Example 2 will be repeated; however, Eq. (8) will be used in place of Eq. (7) and Eq. (18) will replace Eq. (17). Indeed, cumulative numerical integration was performed using Eq. (8), Eq. (12) and Eq. (18) (which will be called Method II). The results are shown in Column I of Table III. Note that these are identical to those in Table II, as they should be, thereby verifying the correctness of Eq. (8) and Eq. (18).

Table III. Implementation of Example 3 using Eq. (8), Eq. (11) with Eq. (18) (Method II) in MS Excel.

	A	B	C	D	E	F	G	H	I	J	K
1	0	0							0	0	0
2	0.1	0.099833							0.00499971	0.00499583	0.00499167
3	0.19	0.188859	0.1	0.09	0.19				0.01799672	0.01799576	0.01798282
4	0.33	0.324043	0.09	0.14	0.23				0.05395332	0.05395766	0.05388596
5	0.4	0.389418	0.14	0.07	0.21				0.07894859	0.07893901	0.07885711
6	0.55	0.522687	0.07	0.15	0.22				0.14746047	0.14747548	0.14726503
7	0.69	0.636537	0.15	0.14	0.29				0.22876683	0.22875399	0.22841073
8	0.74	0.674288	0.14	0.05	0.19				0.26152327	0.26153144	0.26118136
9	0.9	0.783327	0.05	0.16	0.21				0.37839294	0.37839003	0.37779055
10											
11											
12	C3=A2-A1										K2=K1+0.5*(A2-A1)*(B1+B2)
13	D3=A3-A2										J1=1-COS(A1)
14	E3=A3-A1										
15											
16											
17											
18	I3=E3/6*((2-D3/C3)*B1+E3*E3/D3/C3*B2+(2-C3/D3)*B3)+I1										
19	I2=C3/6*((3-C3/E3)*B1+(3+C3*C3/D3/E3+C3/E3)*B2-C3*C3/D3/E3*B3)										
20											
21	*x* values are in column A, *f(x)* in column B.										

Note to create Table III:

(i) $\mathrm{C3} = x_2 - x_1, \mathrm{D3} = x_3 - x_2$ and $\mathrm{E3} = x_3 - x_1$.

(ii) I2 calculates Eq. (8).

(iii) I3 calculates Eq. (6) plus I1.

(iv) Cells C3, D3, E3, and I3 are copied and pasted to partially fill in rows 4 through 9.

(v) K2 calculates Eq. (19).

(vi) K2 is copied and pasted into cells K3:K9.

Conclusion

Two different methods have been presented that allow the numerical cumulative integration of irregularly-spaced data, using Simpson's Rule. These methods are suitable for implementation in spreadsheets. Examples using MS Excel were given.

Appendix

In this appendix, the MATLAB code for the determination of Eq. (8) will be given.

The program requires that Eq. (7) be rewritten as a product given by Eq. (A1).

$$\begin{aligned} I_2 &= \frac{A}{3}\left(x_2^3 - x_1^3\right) + \frac{B}{2}\left(x_2^2 - x_1^2\right) + C\left(x_2 - x_1\right) \\ &= \frac{\left(x_2 - x_1\right)}{6}\left\{2A\left(x_2^2 + x_2 x_1 + x_1^2\right) + 3B\left(x_2 + x_1\right) + 6C\right\}. \end{aligned} \tag{A1}$$

Let $a = x_2 - x_1, b = x_3 - x_2$ and $c = x_3 - x_1$. Then, Eq. (A1) becomes

$$\begin{aligned} I_2 &= \frac{\left(x_2 - x_1\right)}{6}\left\{2A\left[\left(x_1 + a\right)^2 + \left(x_1 + a\right)x_1 + x_1^2\right] + 3B\left(2x_1 + a\right) + 6C\right\} \\ &= \frac{\left(x_2 - x_1\right)}{6}D, \end{aligned} \tag{A2}$$

where $D = 2A\left[\left(x_1 + a\right)^2 + \left(x_1 + a\right)x_1 + x_1^2\right] + 3B\left(2x_1 + a\right) + 6C.$

Furthermore, from Eq. (4),

$$\begin{aligned} A &= \frac{f\left(x_3\right) - f\left(x_2\right)}{bc} - \frac{f\left(x_2\right) - f\left(x_1\right)}{ac}, \\ B &= \frac{f\left(x_3\right) - f\left(x_2\right)}{b} - A\left(2x_1 + a + c\right), \\ C &= f\left(x_1\right) - Ax_1^2 - Bx_1. \end{aligned} \tag{A3}$$

Substituting Eq. (A3) into Eq. (A2) and simplifying produces the desired expression of Eq. (8).

However, the tedium of these final two steps is eased by the MATLAB symbolic math program given below:

```
%simpalgebra2.m Name of program (Comment line)
%Define algebraic symbols (Comment line)
syms f1 f2 f3 x1 a b c
%Now define A, B and C of Eq. (4) (Comment line)
A= (f3-f2)/b/c-(f2-f1)/a/c
B= (f3-f2)/b-A*(2*x1+a+c)
C= f1-A*x1^2-B*x1
%Substitute into D (Comment line)
D=2*A*((x1+a)^2+x1*(x1+a)+x1^2)+3*B*(2*x1+a)+6*C
%Simplify (Comment line)
D=simple(D)
```

```
%Collect like terms (Comment line)
collect(D,f1)
collect(D,f2)
collect(D,f3)
%Make D better readable (Comment line)
pretty(D)
```

Note that the final D as given by this program is a function of $a, b, c, f1, f2$ and $f3$. It is no longer a function of x_1.

Finally, substituting $a = x_2 - x_1, b = x_3 - x_2$ and $c = x_3 - x_1$ into D and using $I_2 = (x_2 - x_1)D/6$ gives the desired Eq. (8).

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

References

[1] Foss, S. D. (1970). A method of exponential curve fitting by numerical integration. *Biometrics*, **26**(4):815-821.

[2] Fresen, J. L., & Juritz, J. M. (1986). A note on Foss's method of obtaining initial estimates for exponential curve fitting by numerical integration. *Biometrics*, **42**(4):821-827.

[3] Jacquelin, J. (2009). Regressions et Equations Integrales. Available:
Scribd : http://www.scribd.com/JJacquelin/documents .

[4] Blake, L.V. (1971). A Simpson's Rule and Fortran Subroutine for Cumulative Numerical Integration Defined by Data Points. NRL Memorandum Report 2231, Naval Research Laboratory. Available:
http://www.dtic.mil/dtic/tr/fulltext/u2/723583.pdf .

[5] El-Hajj, A., & Kabalan, K. Y. (1995). Time domain analysis of linear systems using spreadsheets. *IEEE Transactions on Education*, **38**(4): 317-320.

[6] Cartwright, K. V. (2016). Simpson's Rule Integration with MS Excel and Irregularly-spaced Data. *Journal of Mathematical Science and Mathematics Education*, **11**(2): 34–42.

[7] Dyer, S. A. & Dyer, J. S. (2008). bythenumbers - Numerical integration. *IEEE Instrumentation & Measurement Magazine*, **11**(2): 47-49.