Investigación

The use of Bézier interpolation for numerical integration of uneven data

El uso de la interpolación de Bézier para la integración numérica de datos irregulares

Stefan T Orszulik

Research Journal

Volumen XV, pp. 003–012, ISSN 2174-0410
Recepción: 22 Abri'24; Aceptación: 27 May'24

1 de abril de 2025

Abstract

One of the challenges in mathematics is to find the area under a curve when the usual methods of integration are not feasible. There are various approaches to deal with this problem, most of them relying on Newton-Cotes formulas, but they have some drawbacks and errors. This article presents a different method that uses Bézier interpolation to approximate the curve and calculate the area. This method has the advantage of being accurate and general, meaning that it can handle any kind of curve, even those that come from experimental data that may be irregular or noisy.

Keywords: Bézier interpolation, Simpson's rule, trapezoidal rule, numerical integration, estimation.

Resumen

Uno de los desafíos en matemáticas es encontrar el área bajo una curva cuando los métodos habituales de integración no son factibles. Existen varios enfoques para abordar este problema, la mayoría de ellos basándose en fórmulas de Newton-Cotes, pero tienen algunos inconvenientes y errores. Este artículo presenta un método diferente que utiliza la interpolación de Bézier para aproximar la curva y calcular el área. Este método tiene la ventaja de ser preciso y general, lo que significa que puede manejar cualquier tipo de curva, incluso aquellas que provienen de datos experimentales que pueden ser irregulares o ruidosos.

Palabras Clave: interpolación de Bézier, la regla de Simpson, regla trapezoidal, integración numérica, estimación.

1. Introduction

One way to approximate the integral of a function that is hard or impossible to integrate analytically is to use the trapezoidal rule [1]. This technique divides the area under the curve into several trapezoids (see Figure 1), and adds up the area of each trapezoid to get an estimate of the total area. The main benefit of this technique over other methods is its simplicity, especially when implemented in a spreadsheet, since the formulas can be easily copied and pasted, making it a popular alternative to direct integration; however, the technique also has some inherent errors that restrict its applicability.

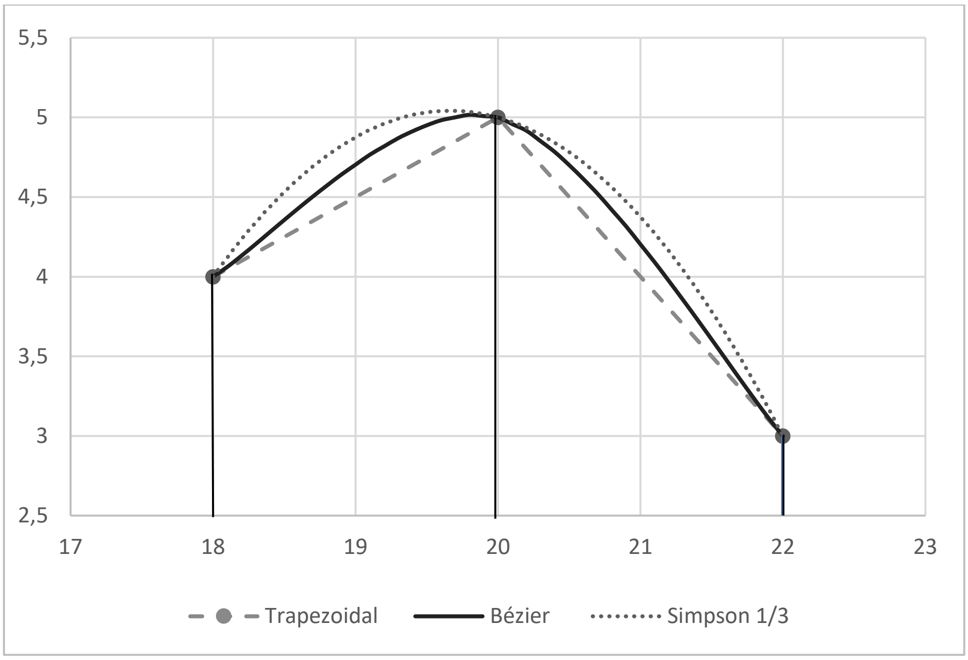

Figure 1. Bézier curve together with Trapezoidal Rule and Simpson's Rule for two segments.

In the standard trapezoidal method, the area under the curve (x = a to b) is:

$$I = (b - a) * [f(a) + f(b)]/2 \quad (1)$$

As can be seen in Figure 1, there is a section between the curve and the trapezoidal shape that is not included in the calculation, leading to error.

Simpson's rule overcomes much of this error by fitting two sections of the curve to a quadratic fit.

$$I = \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] \quad (2)$$

Simpson's rule is a numerical method for approximating the area under a curve that is smooth and continuous. It works well when the curve can be modelled by a quadratic function, or a polynomial of degree two. However, it may not be very accurate when the curve has a different shape, such as a sine or cosine function. To apply Simpson's rule to a curve that is divided into several subintervals, we can use the formula in Eq. (3), which sums up the areas of each subinterval using Simpson's rule.

$$I = \frac{(b-a)}{3n}[f(a) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \cdots + 4f(x_{n-1}) + f(b)] \quad (3)$$

The method requires an even number of equally spaced intervals (n is even). When the function of the curve is known it is easy to arrange for an even number of intervals. Increasing the number of intervals increases the accuracy of both the Trapezoidal Rule and Simpson's 1/3 Rule (Eq. (3)); therefore, when there is a function describing the curve then it is relatively easy to increase the number of intervals, particularly when using the copy-and-paste facility in a spreadsheet.

There are other Newton–Cotes formulas such as Boole's Rule and Weddle's Rule which use higher number of intervals (n) but these have inherent drawbacks; firstly, they can only be applied to larger values of n (or their multiples); secondly, use of larger values of n can suffer from Runge's phenomenon where the errors grow significantly for large n. There have been several proposals for combining different Newton–Cotes formulas for various values of n which have been shown to give good results [2, 3], but each value of n has its own combination and therefore cannot be easily applied universally. In addition, these algorithms can only be applied to equally spaced intervals.

In practice there are many occasions when there is no function associated with the curve, and indeed where the data is not evenly distributed, as is the case with much experimental data. The trapezoidal rule can still be applied for data that are unevenly spaced, but this is not the case with Simpson's 1/3 Rule which usually requires equally spaced intervals. Furthermore, when the data is obtained experimentally and not easily associated with a mathematical function, it is not possible to increase the number of intervals.

More recently there have been reports of a modifies Simpson's 1/3 Rule that is applicable to irregularly spaced data [4, 5]. Particularly useful is Cartright's article [4] since it is easily applied using MS Excel. The equation for Simpson's 1/3 Rule with irregular spaced intervals is given in Eq. (4).

$$I = \frac{x_3-x_1}{6}\left[\left\{2 - \frac{x_3-x_2}{x_2-x_1}\right\} f(x_1) + \frac{(x_3-x_1)^2}{(x_3-x_2)(x_2-x_1)} f(x_2) + \left\{2 - \frac{x_2-x_1}{x_3-x_2}\right\} f(x_3)\right] \quad (4)$$

Eq. (4) is for two subintervals. Once set up in Excel it is very straightforward to copy the equation to providing a numerical integration over the entire dataset.

Figure 1 shows a Bézier curve, which is a type of parametric curve that goes through all the datapoints using control points. The first and last control points are at the beginning and end of the curve. Other control points are placed such that the resultant parametric curve passes through the data points. The following shows a cubic Bézier function:

$$B(t) = (1-t)^3 P_0 + 3(1-t)^2 t P_1 + 3(1-t) t^2 P_2 + t^3 P_3 \quad (5)$$

where the parameter, t, has a value between 0 and 1, and P_i are the control points.

You can find more information about Bézier curves and how they are used in the literature, for example, by Hosaka [6].

This article presents a method based on Bézier curve fitting that could be better than other numerical integration methods in many cases, especially when applied to unevenly spaced data.

2. Method

One way to calculate the area under a curve is to use integration, which can be done analytically if the curve has a known mathematical expression. However, many times the curve is obtained from experimental measurements that do not have a simple formula. In these cases, numerical methods such as the trapezoidal rule or Simpson's rule can be used to approximate the integral using the data points. To illustrate how these methods work, we will apply them to some curves that have known functions and compare the results with the exact values. Then we will show how a new method based on Bézier curves can be used to estimate the integral of experimental data and how it compares visually with the other methods.

3. Results

3.1 Examples with known integrals

The article by Ullah [3] compared several integration methods using equations with known integrands. One of these (Eq. (6)) is examined here, comparing the Trapezoidal and Simpson's 1/3 Rules with the proposed Bézier method.

$$\int_0^1 \sqrt{1-x^2}dx \tag{6}$$

Table 1 is a portion of a spreadsheet showing the proposed integration method using Bézier interpolation.

Table 1. Numerical Integration using Bézier interpolation (first 13 of 100 intervals shown for Bézier interpolation data).

		Bézier Interpolation		
Known X values	*Known Y values*	*X*	*Interpolated Y*	*Trapezoidal Rule*
0	1	0	1	
0.0769	0.9970	0.01	0.9997	0.009999
0.1538	0.9881	0.02	0.9995	0.009996
0.2308	0.9730	0.03	0.9992	0.009993
0.3077	0.9515	0.04	0.9989	0.009990
0.3846	0.9231	0.05	0.9985	0.009987
0.4615	0.8871	0.06	0.9981	0.009983
0.5385	0.8427	0.07	0.9975	0.009978
0.6154	0.7882	0.08	0.9968	0.009971
0.6923	0.7216	0.09	0.9959	0.009964

0.7692	0.6390	0.1	0.9950	0.009955
0.8462	0.5329	0.11	0.9939	0.009945
0.9231	0.3846	0.12	0.9928	0.009933
1	0.0000	0.13	0.9915	0.009921

The data in Table 1 are as follows:

- Columns 1 and 2 are 13 equally spaced segments based on Eq. (6).
- Columns 3 and 4 are the Bézier interpolation based on the data in Columns 1 and 2. Thus column 3 consists of 100 equally spaced segments running from 0 to 1 (the integral limits) of which only the first of thirteen segments are shown in Table 1.
- The syntax for column 4 is "Bezier(KnownXValues, KnownYValues, XToInterpolate)". This MS Excel function is available from reference by Lépissier [7].
- Column 5 is the Trapezoidal Rule integral of the Bézier interpolation data. The sum of this column (not shown) is the numerical integration using the Bézier interpolation; this amounts to 0.781523, very close to the exact value.

Table 2 below presents some examples from Ullah [3] along with this result. The table shows that Bézier interpolation gives the most accurate results compared to Trapezoidal Rule and Simpson's 1/3 Rule in most cases. The only exception is the last example, which is still very close to the true value.

It is instructive to examine the last example more closely. The curve between 0 and 1 has a concave and convex region. When examined separately (before and after the point of inflection) the Bézier method proves to be more accurate than the other methods, including Simpson's 1/3 Rule. However, Simpson's Rule overestimates the integral on one section of the curve and underestimates it on the other section. Taken together, these partly cancel themselves leading to an apparent lower overall error in this case.

The data in Table 2 indicate that Bézier interpolation is a reliable method for numerical integration. Moreover, as Table 1 demonstrates, the method is easy to implement in a spreadsheet and can handle any dataset; it does not require a fixed number of intervals like Simpson's 1/3 Rule and most other Newton–Cotes formulas.

Table 2. Comparison of numerical integration methods for n = 13

Integral	Exact Value	Trapezoidal	Simpson's 1/3 + Trapezoidal*	Bézier + Trapezoidal**
$\int_0^1 \sqrt{1-x^2}dx$	0.78539816	0.779140612	0.780272	0.781523
	Error	0.006258	0.005126	0.003875
$\int_0^2 (e^{x^2}-1)dx$	14.45262777	14.8797	14.6677	14.56846402
	Error	0.42707	0.21507	0.11584
	3.26107456	3.259873	3.260673	3.261264001

$\int_0^2 \sqrt{1+3\sin^2 x}\,dx$	Error	0.001202	0.000402	0.00019
$\int_0^1 (1+e^{-x}\cos(4x))dx$	1.00745963	1.00862398	1.007536	1.007292649
	Error	0.00116	0.000076	0.000167

* Simpson's 1/3 rule for n = 12 plus Trapezoidal Rule for the final segment.

** Bézier interpolation generating a curve with n = *100* followed by Trapezoidal Rule.

3.2 Examples with unknown integrals and unequal intervals

3.2.1 Example 1

Figure 2 is a chart containing data that does not have a function associated with it, and indeed where the data are not equally distributed. Included in the chart is the Bézier curve fitting the datapoints as well as the trapezoidal rule. Figure 3 shows the same dataset but includes the Simpson's 1/3 Rule alongside the Bézier fit.

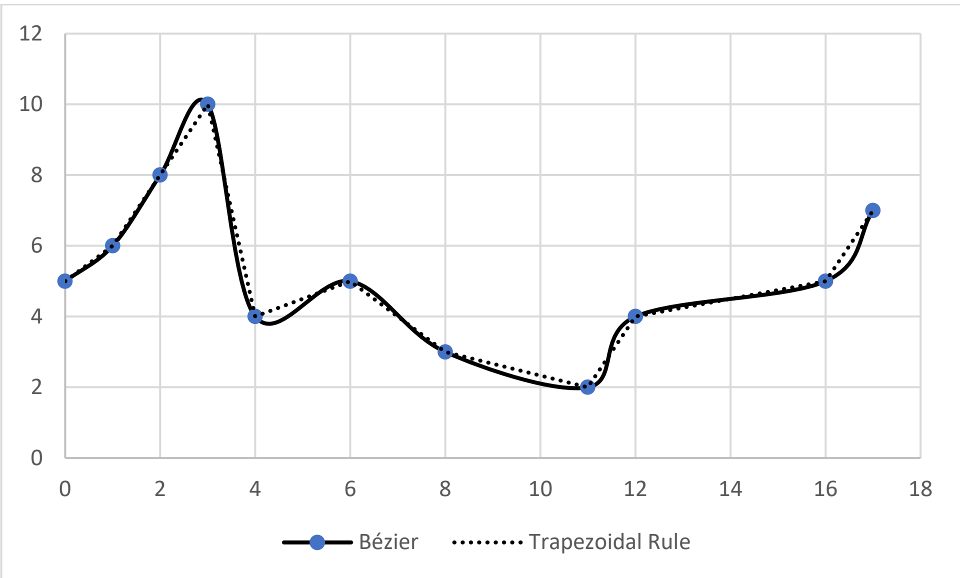

Figure 2. Chart demonstrating a Bézier fit and Trapezoidal rule for datapoints with unequal intervals.

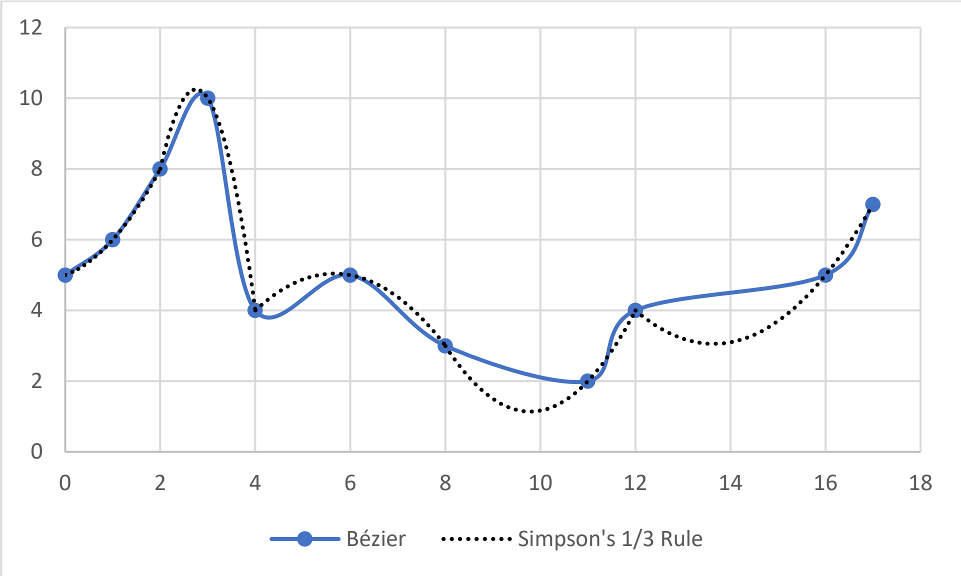

Figure 3. Chart demonstrating a Bézier fit and Simpson’s Rule for datapoints with unequal intervals.

The data shown in Figures 2 and 3 are examples of what might be obtained from experiments; they do not follow any mathematical formula, and they have irregular intervals. To calculate the area under such curves, the Trapezoidal Rule or Simpson’s Rule for unequal intervals are possible options. However, there is no objective criterion to choose between them, so the researcher has to rely on their own judgement. Looking at Figures 2 and 3, it seems that Bézier interpolation would give a better approximation than the other methods. Moreover, the Bézier method is easy to implement in a spreadsheet, can handle unequal intervals, and does not require an even number of intervals like Simpson’s 1/3 Rule.

Figures 2 and 3 have both concave and convex parts, so the errors of the Trapezoidal and Simpson’s Rules are partly offset, as seen on the graphs. The values of the integrals obtained by the three methods in this example are given in Table 3.

Table 3. Comparison of Integrals generated by the three numerical integration methods for n = 13

Trapezoidal Rule	Simpson's Rule	Bézier Interpolation + Trapezoidal*
80.00	75.65	79.51

* Bézier interpolation generating a curve with $n = 170$ followed by Trapezoidal Rule.

3.2.2 Example 2

Figure 4 illustrates the results of applying three different numerical integration methods to a set of experimental data from Google Documents [8]. The data points are not evenly distributed along the x-axis, which poses a challenge for estimating the area under the curve. The three methods are the Trapezoidal Rule, Simpson's 1/3 Rule, and the Bézier curve fitting.

The values of the integrals obtained by each method are shown in Table 4. The figure suggests that the Bézier curve provides a smoother and more accurate approximation of the original data than the other two methods.

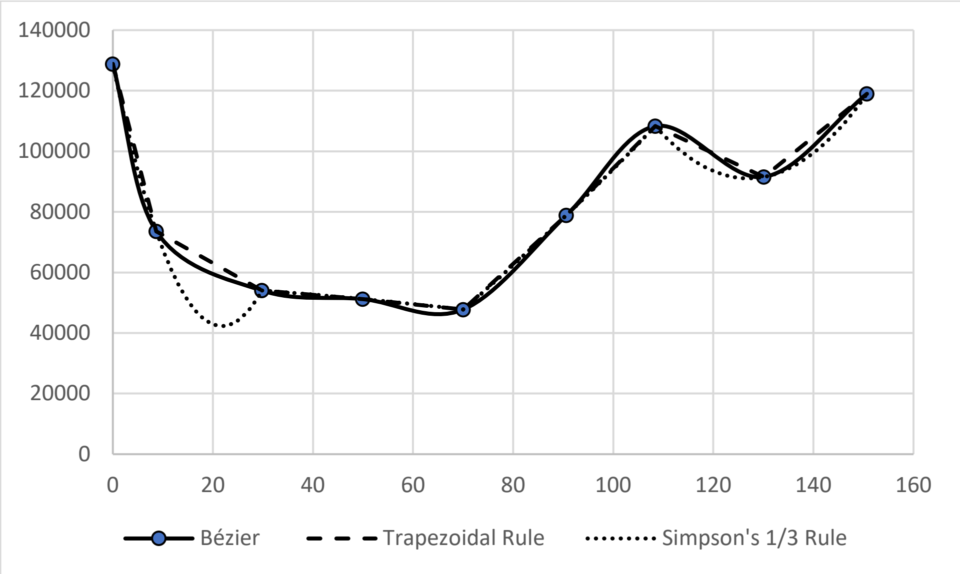

Figure 4. Chart demonstrating a Bézier fit, Simpson's Rule and Trapezoidal Rule for datapoints with unequal intervals.

Table 4. Comparison of Integrals methods.

Trapezoidal Rule	Simpson's Rule	Bézier Interpolation + Trapezoidal*
11585086	11115587	11439809

* Bézier interpolation generating a curve with $n = 100$ followed by Trapezoidal Rule.

4. Conclusions

This article presents a Bézier method for numerical integration, which is a process of finding the area under a curve when analytical integration is not feasible. Unlike many other numerical methods, such as those based on Newton-Cotes formulas, the Bézier method does not have limitations on the number or spacing of the intervals. It can handle any dataset, even if the intervals are uneven or irregular. The method is also accurate, as it has been shown to have a low error rate compared to other methods. Therefore, the Bézier method could be a useful tool for numerical integration, especially for experimental data that may not follow a regular pattern.

References

[1] CHAPRA, S.C., CANALE. R.P., 2015. *Numerical methods for engineers. 7th ed.*, pp. 605-615.

[2] BHATTI, A.A, CHANDIO, M.S., MEMON, R.A., SHAIKH, M.M., 2019. *A Modified Algorithm for Reduction of Error in Combined Numerical Integration,* Sindh Univ. Res. Jour. Sci. Ser. 51(04), 745-750, doi.org/10.26692/sujo/2019.04.

[3] ULLAH, M., 2015. *Numerical Integration and a Proposed Rule,* Amer J Engineering Research, 4(9), pp-120-123

[4] CARTWRIGHT, K.V., 2016. S*impson's Rule Integration with MS Excel and Irregularly-spaced Data,* J Math Sci & Math Ed. 11(2), 34.

[5] SINGH, A.K., THORPE, G.R., 2023, *Simpson's 1/3 Rule of Integration for Unequal Divisions of Integration Domain,* https://www.researchgate.net/publication/265499536_Simpson%27s_13-rule_of_integration_for_unequal_divisions_of_integration_domainA.K. SINGH and G. R. THORPE. (Accessed 12-12-2023).

[6] HOSAKA, M., 1992. *Bézier Curves and Control Points. In: Modeling of Curves and Surfaces in CAD/CAM. Computer Graphics — Systems and Applications.* Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-76598-8_9

[7] LÉPISSIER A, Coding - Dr. Alice Lépissier (alicelepissier.com), accessed 02/01/2024

[8] https://docs.google.com/spreadsheets/d/1qZZtCFRBzGmO8f6ZP1toFEagdAkTaejuj1CPKYpvX_8/edit#gid=0. (Accessed 01-01-2024).

About the author:
Name: Stefan T Orszulik
Email: s.orszulik@ntlworld.com
Institution: Dr Orszulik obtained a BSc and PhD from Royal Holloway College, University of London, followed by postdoctoral research at Strathclyde University, Glasgow.
Current address: 6, The Kestrels, Grove, Wantage, Oxfordshire OX12 0QA, UK.

Orcid: orcid.org/my-orcid?orcid=0000-0003-3176-4911