Estimating the error in the numerical integration: Trapezoidal rule

All numerical integration algorithms inherently contain an element of uncertainty associated with the method used to interpolate between points. If the points sampled are exact representations of the function being integrated, higher-order interpolation methods will yield more accurate results (assuming that the points are spaced so as to capture the highest non-zero derivatives). If, however, the points sampled do not fully capture the underlying function, any uncertainty in the values of these points is amplified by the use of higher-order methods. The Trapezoidal rule is therefore preferred over high-order methods such as Simpson's method when dealing with uncertain data.

When performing a numerical integration using the Trapezoidal rule the total error due to the uncertainty in the points sampled ($y_i \pm \sigma_{yi}$) can be determined analytically by applying the general Gaussian error propagation formula.[1] If the errors can be treated as independent and normally distributed, this yields Eq. 1.

$$Error = \frac{1}{2}\sqrt{(x_1 - x_0)^2\sigma_{y_0}^2 + \sum_{i=1}^{N-1}(x_{i+1} - x_{i-1})^2\sigma_{y_i}^2 + (x_N - x_{N-1})^2\sigma_{y_N}^2} \quad (1)$$

Where an arbitrary function is represented by points (x_i, y_i) with uncertainty in the y_i values given by σ_{yi}.

Another source of error associated with numerical integration methods is due to the implicit assumption that sufficient points are present to capture the highest non-zero derivative of the underlying function being integrated. However, in many cases numerical integration is used precisely because the form of the underlying function is unknown. This source of error, often referred to as truncation error, is more difficult to quantify than that due to the uncertainties in the points. Using a Taylor series expansion, it can be shown that in the case of the Trapezoidal rule the truncation error for the interval between two points a and b can be estimated as:

$$Error \approx -\frac{(b-a)^3}{12}f''(\xi) \quad (2)$$

where $f''(\xi)$ is the second derivative over the interval [a, b]. The second derivative over this interval can be calculated numerically by finite difference methods and the 2nd order truncation error estimated for each interval. The uncertainty when using the Trapezoidal rule can therefore be estimated as the absolute value of the sum of Eq. 2 applied to each interval [x_i, x_{i+1}], and the point-uncertainty propagation given by Eq. 1. Summing the truncation error for each interval (i.e. not taking their absolute value first) allows for the cancelation of errors for functions containing both concave and convex regions. However, uncertainty in the predicted truncation error can lead to the inappropriate cancellation of errors resulting in a significant underestimation of the total error. Two additional steps can be taken to improve the robustness of the error estimate: 1) truncation errors were calculated using both forward and backward finite difference differentiation methods and the method producing the largest error was used; and, 2) the absolute value of the maximum interval truncation error was added to the final error estimate. Thus the overall error estimate is given by Eq. 3.

$$
\begin{aligned}
Error = &\frac{1}{2}\sqrt{(x_1 - x_0)^2\sigma_{y_0}^2 + \sum_{i=1}^{N-1}(x_{i+1}x_{i-1})^2\sigma_{y_i}^2 + (x_N - x_{N-1})^2\sigma_{y_N}^2} \\
&+ \max\left(\left|\sum_{i=1}^{N-1} Er_i^{\text{forward}}\right|, \left|\sum_{i=1}^{N-1} Er_i^{\text{backward}}\right|\right) \\
&+ \max\left(\max\{|Er_i^{\text{forward}}|\}_{i=0}^{N-1}, \max\{|Er_i^{\text{backward}}|\}_{i=0}^{N-1}\right)
\end{aligned} \tag{3}
$$

Where Eri^{forward} and Eri^{backward} are the interval truncation errors calculated using Eq. 2 where $f''(\xi)$ over the interval $[x_i, x_{i+1}]$ is estimated using the forward and backward finite difference differentiation methods respectively.

Note that the truncation error (given by Eq. 2) is not entirely independent of the uncertainty in the error predicted at each point (Eq. 1) since the points themselves are used to estimate $f''(\xi)$ over the interval $[x_i, x_{i+1}]$. While it is in principle possible to propagate uncertainties considering the various correlations using the Gaussian error propagation formula, the practical utility of considering such higher order effects is unclear.

For more details see Stroet[2].

References

1. Ku, H. H., Notes on the use of propagation of error formulas. *J. Res. Natl. Bur. Stand.* **1966**, *70.*
2. Stroet, M. Improving the accuracy of molecular dynamics simulations: parameterisation of interaction potentials for small molecules. The University of Queensland, **2018**, doi:10.14264/uql.2018.432.