Title: To what extent can the Error between the Deterministic Newton Cotes and Stochastic Monte Carlo Quadrature Rules for Numerical Integration be meaningfully quantified?

Research Question: What is a rigorous quantification of the error in both the deterministic Newton Cotes and stochastic Monte Carlo quadrature rules?

Subject: Mathematics

Table of Contents

0 Introduction

It has been three centuries since the birth of Calculus and it is integral to acknowledge that it has come a long way from when it was "brought to the threshold of existence" by Archimedes of Syracuse through the development of "protointegration" methods enabling him to calculate areas under curves such as the parabola (Dunham, W. ,2018, p.1). It is a field of inquiry with "a rich history and a rich prehistory" (Dunham, W. ,2018, p.1). By means of a few prolific individuals such as Isaac Newton, Gottfried Leibniz, Augustine - Louis Cauchy and Henri Lebesgue, the toolkit of this field of endeavour has been vastly expanded and through the passage of time, its rigorisation has bestowed it with eternal prestige and prominence in mathematics. However, the essence of Calculus when it was conceived, was to determine the area under a curve that need not be linear, whence the process of integration was born. The process of finding analytic, namely, exact solutions to such integrals, albeit harbouring a multitude of powerful techniques, has inherent limitations, as expressed by Liouville's theorem; it places restrictions "on antiderivatives that can be expressed as elementary functions", namely functions such as powers, roots, exponentials, logarithms, trigonometric and hyperbolic functions. ("Liouville's theorem", 2019).

The above necessitates the development of methods that will be able to yield numerical values as solutions to definite integrals that either have no elementary antiderivative, or are too cumbersome to find analytic solutions for. A preliminary investigation into a small portion of the numerical methods thereof is thus pertinent due to numerical integration seeing a manifold of applications from physics and engineering to probability and statistics. This has been facilitated by the development of powerful computing resources that implement the above methods to great effect.

In this investigation, we shall initially motivate and derive the formula for polynomial interpolation as an approximation to a function of interest. After deriving a formula for the error, we shall consider the **Trapezium Rule** in both its single and composite case. Generalising, we will integrate the expression for the **polynomial interpolation error** over an interval to obtain the **Newton Cotes quadrature rules**. Thereafter, we will investigate a **stochastic quadrature rule**, namely, the **Monte Carlo method**. We will show some preliminary results involving continuous random variables (such as the linearity of expectation and variance for independent random variables) and then derive a "probabilistic" estimate of the error and conclude by comparing the above methods and how they perform computationally through the use of a representative example (see Section 4).

1 Polynomial Interpolation

It is important that we begin our investigation with some motivation for the development of the forthcoming formalism. Calculus has powerful results for functions that behave sufficiently 'nicely'. That is, they are continuous and differentiable, which is quite restrictive and comes in conflict with many real-life situations where full knowledge of the modelling function may be limited to a discrete set of data points. Our problem can thus be restated to being "able to construct continuous functions based on discrete data." (Jim Lambers, 2010-11). A reasonable such class of functions that would fit the data would be polynomials as they are elementary and easy to work with. Even if the modelling function is known which is indeed what we shall assume henceforth, the above can prove to be a useful simplifying assumption, especially in the field of numerical integration. I thus begin by constructing explicitly the polynomial - which will turn out to be unique - that will be needed to interpolate a set of points, which can either be real data points or points on a given modeling function.

1.1 Lagrange Interpolation

Theorem 1.1 *Let* $S = \{x_0, x_1, \ldots, x_n\}$ *be a set of distinct real numbers and let* $f(x)$ *be a function defined on a domain including S, then there exists a unique polynomial* $p_n(x)$ *of degree* n *or less, such that* $\forall x_i \in S,\ p_n(x_i) = f(x_i)$ (Yan-Bin Jia, 2017). The polynomial is defined by

$$p_n(x) = \sum_{j=0}^{n} f(x_j) L_{n,j},$$

where

$$L_{n,j} = \prod_{k=0,\, k\neq j}^{n} \frac{x-x_k}{x_j-x_k}.$$

Proof

For uniqueness, it is enough to show that for two different polynomials p, q with degree at most n, if they satisfy the interpolating property delineated in the statement of the theorem, then $p \equiv q$. We define the polynomial $p - q$ of degree at most n with roots at each of the x_i. Thus, the above polynomial can be expressed as

$$(p - q)(x) = c \prod_{i=0}^{n} (x - x_i)A(x)$$

for some $c \in \mathbb{R}$ and $A(x)$is a polynomial in x accounting for the other potential roots of $(p - q)(x)$.By comparing the degrees of the left hand side and the right hand side, we realise that the LHS has a degree of at most n, whilst the RHS has a degree of at least $n + 1$; in order to avoid a contradiction only when c = 0 and thus $p \equiv q$.

(Yan-Bin Jia, 2017)

Why is this result useful? It guarantees the existence and uniqueness of such a polynomial approximator. Now, the central object of our focus is to quantify the difference between this polynomial and the function we wish to interpolate. Hence, making a step in that direction, we are going to prove an important result that will enable us to make progress with regards to the quantification of the error of the approximator polynomial.

1.2 Rolle's Theorem for Higher Derivatives

Lemma 1.1 *(Generalised Rolle's Theorem) Let* f *be continuous on the closed interval* $[a, b]$*and n times differentiable in the open interval* (a, b)*. If* f *is zero at the* $n + 1$ *distinct points* $x_0 < x_1 ... < x_n$ *in* $[a, b]$*, then there exists a number* $c \in (a, b)$ *such that* $f^{(n)}(c) = 0$. (Jim Lambers, 2010-11)

Proof

I shall use the principle of mathematical induction.

If $n = 1$, then the result is a restatement of the standard Rolle's Theorem (See Weisstein, E. (2019)) .

For the sake of clarity, we state the inductive step: "*Let* f *be continuous on the closed interval* $[a, b]$*and n times differentiable in the open interval* (a, b)*. If* f *is zero at the* $n + 1$ *distinct points* $x_0 < x_1 ... < x_n$ *in* $[a, b]$*, then there exists a number* c *in* (a, b) *such that* $f^{(n)}(c) = 0$ "(Alexander, 2014).

We fix $n > 1$ and assume the result for n. We now have to show that this is true for the case $n + 1$, namely, when *"*f *be continuous on the closed interval* $[a, b]$*and n+1 times differentiable in the open interval* (a, b)*. If* f

is zero at the $n + 2$ *distinct points* $x_0 < x_1 ... < x_{n+1}$ *in* $[a, b]$, *then there exists a number* c *in* (a, b) *such that* $f^{(n+1)}(c) = 0$ ”(Alexander, 2014).

We apply Rolle’s Theorem on each of the consecutive pairs of points x_{i-1}, x_i, $1 \leq i \leq n+1$ and obtain a set of $n + 1$ numbers y_i, such that $f'(y_i) = 0$.
Now, we make use of the induction hypothesis as $f'$is n times differentiable, where f is replaced with $f'$and the n points are the $\{y_i\}$indexed by the same i above.

Thus, we get that there is a $\xi \in (a, b)$ such that $f'^{(n)}(\xi) = 0 = f^{(n+1)}(\xi)$, which finishes the proof.

(Alexander, 2014)

Now we have the necessary tools to give proof of the interpolation error for polynomials, which is the central result of this section.

1.3 Interpolation Error

Theorem 1.2 *(Interpolation Error) If* f *is* $n+1$ *times continuously differentiable on* (a, b), *and* $p_n(x)$ *is the unique polynomial of degree* n *that interpolates* f *at the* $n+1$ *distinct points* $x_0, x_1, ..., x_n$, *in* $[a, b]$, *then for each* $x \in [a, b]$,

$$f(x) - p_n(x) = \prod_{j=0}^{n} (x - x_j) \frac{f^{(n+1)}(\xi_x)}{(n+1)!}$$

for some $\xi_x \in (a, b)$.

(Bruce Edwards)
Proof

We begin by fixing an arbitrary $x \in [a, b]$ not one of the interpolation points, then we define

$$\varphi(t) = f(t) - p_n(t) - \frac{f(x)-p_n(x)}{\pi_{n+1}(x)}\pi_{n+1}(t),$$

where

$$\pi_{n+1}(x) = \prod_{i=0}^{n}(x - x_i),$$

is a polynomial of degree $n + 1$.

From the definition of $\varphi(t)$, $\varphi(x_i) = 0$, $0 \le i \le n$ or when $t = x$.Thus it has $n + 2$ roots, and is $n + 1$ times differentiable in (a, b). Using lemma 1.1, there exists a $\xi \in (a, b)$ such that $\varphi^{(n+1)}(\xi) = 0$.

But,

$$\varphi^{(n+1)}(t) = \frac{d^{(n+1)}}{dt^{(n+1)}}\left(f(t) - p_n(t) - \frac{f(x)-p_n(x)}{\pi_{n+1}(x)}\pi_{n+1}(t)\right)$$

$$= \frac{d^{(n+1)}}{dt^{(n+1)}}(f(t)) - \frac{d^{(n+1)}}{dt^{(n+1)}}(p_n(t)) - \frac{d^{(n+1)}}{dt^{(n+1)}}\left(\frac{f(x)-p_n(x)}{\pi_{n+1}(x)}\pi_{n+1}(t)\right)$$

$$= f^{(n+1)}(t) - p_n^{(n+1)}(t) - \frac{f(x)-p_n(x)}{\pi_{n+1}(x)}\frac{d^{(n+1)}}{dt^{(n+1)}}\left(\prod_{i=0}^{n}(t - x_i)\right).$$

Thus,

$$\varphi^{(n+1)}(t) = f^{(n+1)}(t) - 0 - \frac{f(x)-p_n(x)}{\pi_{n+1}(x)}(n + 1)!$$

and hence,

$$\varphi^{(n+1)}(\xi) = 0 = f^{(n+1)}(\xi) - 0 - \frac{f(x)-p_n(x)}{\pi_{n+1}(x)}(n+1)!.$$

Rearranging gives the required solution (Bruce Edwards). ▨

Remark 1.2

One may informally observe that

$$\frac{(n+1)![f(x) - p_n(x)]}{\prod_{j=0}^{n}(x-x_j)}$$

is continuous assuming that x is not one of the interpolating points (to avoid division by zero); this leads one to deduce that $f^{(n+1)}(\xi_x)$is a continuous function. This will be elaborated on in a more rigorous manner later in Section 2.3, when it will be necessary to achieve complete rigour.

It is important to note that we can give an upper bound on $f^{(n+1)}(x)$ on $[a, b]$ and hence $|f^{(n+1)}(x)|$ as the former is continuous on the closed interval $[a, b]$(Extreme Value Theorem, Wolfram MathWorld).

Corollary 1.1 *Let* .

$$\max_{x \in [a,b]} |f^{(n+1)}(x)| = M$$

Then,

$$|f(x) - p_n(x)| \le M \left| \prod_{j=0}^{n} (x - x_j) \right|.$$

Proof

The proof follows immediately from substituting $|f^{(n+1)}(x)| \leqslant M$ into the absolute value of the result of Theorem 1.2. ▨

The corollary is useful as it provides us with an upper bound for the absolute value of the error of the interpolating polynomial. Yet, concerning the research question, this in conjunction with the error form in

Theorem 1.2, may spell issues regarding an **exact and computable** determination of the error as the only meaningful recourse would be to bound the error.

One caveat of the above method is that for large n, the polynomial interpolation may become problematic due to the highly oscillatory nature of the approximating polynomial *(Bruce Edwards).* This is a problem known as overfitting which is frequently encountered in areas of inquiry such as machine learning, where the predictability of an algorithm may fall after a critical number of data points (or point on a function) are collected (see figure 1).

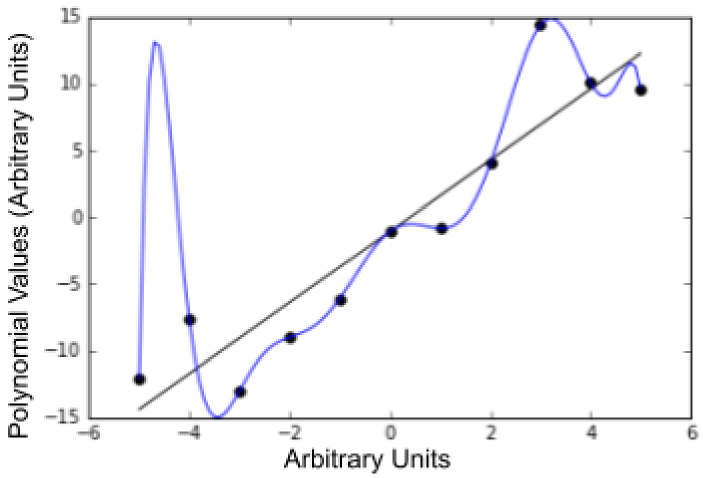


Figure1: Data that is roughly linear yields a wildly oscillatory polynomial interpolant - if many points are to be interpolated - that will give rapidly divergent extrapolations compared to the linear fit. (Wikipedia, Overfitting)

2 Deterministic Numerical Integration

Now that we are more knowledgeable of the error involved in interpolating a function using a polynomial, we may apply it to the study of integration by numerical means, where we produce an estimate of a given definite Riemann integral, whose exact solution may be too hard or downright impossible to obtain by analytic means.

In any case, we will proceed with the proof of a result which is going to prove essential in the derivation of the error estimate for the composite Trapezium Quadrature Rule; this will be achieved by simplifying the integral expression for the estimate thereof. More specifically, the following result is of significance because it will enable us to 'pull' out a term of an integral expression for the error of the numerical integration quadrature leading to the integration of a quadratic function.

2. 1 Mean Value Theorem for Integrals

Theorem 2.1 *(First Mean Value Theorem for Integrals) Suppose that* u *is continuous on* $[a, b]$ *and* v *is integrable and nonnegative on* $[a, b]$. *Then,*

$$\int_a^b u(x)v(x)dx = u(c)\int_a^b v(x)dx$$

for some $c \in (a, b)$(William Trench, 2013).

Proof

Since $u(x)$ is continuous on a closed interval, let $m = min_{x\in[a,b]} u(x)$ and $M = max_{x\in[a,b]} u(x)$. Then, $m \le u(x) \le M$ and since $v(x) \ge 0$, the inequality $m.v(x) \le u(x)v(x) \le M.v(x)$ is preserved. We take integrals of both sides and we thus get,

$$m\int_a^b v(x)dx \le \int_a^b u(x)v(x)dx \le M\int_a^b v(x)dx.$$

We further assume that $\int_a^b v(x)dx \neq 0$because if $\int_a^b v(x)dx = 0$ then from the above inequality, the result becomes true for all $c \in (a, b)$.

Thus,

$$m \le \frac{\int_a^b u(x)v(x)dx}{\int_a^b v(x)dx} \le M$$

after dividing by

$$\int_a^b v(x)dx$$

and using the Intermediate Value Theorem (see Intermediate Value Theorem. (n.d.)), there exists a $c \in (a, b)$ such that

$$\frac{\int_a^b u(x)v(x)dx}{\int_a^b v(x)dx} = u(c)$$

and this finishes the proof. ▨
(William Trench, 2013)

Remark 2.1 *We note that the function $v(x)$ can equally be always negative in the interval. The proof is identical as the negativity flips the order of the inequality $mv(x) \leq u(x)v(x) \leq Mv(x)$ which is reversed again by the division of* $\int_a^b v(x)dx$*.*The essential idea is that $v(x)$must exhibit **no sign change** over $[a, b]$.

2.2 Trapezoidal Rule Error for the Composite Case

Moving on, we shall derive the quadrature rule for the composite Trapezium Rule. The scenario is illustrated in figure 2. Here an interval [a,b] is partitioned into N smaller intervals of equal length h = Δx where the Trapezium Rule will be applied to each of the subintervals comprised of adjacent points to obtain the approximate quadrature and error.

In particular, the above interval [a,b] is partitioned into the N subintervals $[x_k, x_{k+1}]$,where the points are equally spaced; thus $x_k$is of the form $a + kh$, where $h = \frac{(b-a)}{N}$. Using figure two as our visual guide, we realise that the estimated quadrature is equal to the sum of the areas of the trapezoids that lie above the given subintervals. This can be made rigorous using the Lagrange formalism that we developed in Section 1. In any case, the estimate for the Riemann Integral $\int_a^b f(x)dx$ is

$$T(f, h) = \frac{h}{2}(f(x_0) + 2\sum_{i=1}^{N-1} f(x_i) + f(x_N).$$

Now for a quantification of the error of the above estimated quadrature.

Theorem 2.1 ***(Error Analysis Composite Trapezoid Rule)*** *If f is has a continuous second derivative on [a,b], then there exists* $\eta \in (a, b)$ *such that*

$$\int_a^b f(x)dx - T(f, h) = \frac{h^2 (a-b) f^{(2)}(\eta)}{12}.$$

(Numerical Integrations, 2018)

Proof

We begin by showing the following:

$$f(x) - P_k(x) = (x - x_k)(x - x_{k+1})\omega_k(x).$$

Where $P_k(x)$is the interpolating line at each subinterval $[x_k, x_{k+1}]$and $\omega_k(x)$ is a **continuous** function on $[x_k, x_{k+1}]$; we will proceed by constructing a suitable $\omega_k(x)$. This is done to ensure the existence of the Riemann integral

$$\int_{x_k}^{x_{k+1}} \omega_k(x)(x - x_k)(x - x_{k+1})\, dx.$$

Let us consider the function

$$\psi_k(x) = \frac{f(x)-P_k(x)}{(x-x_k)(x-x_{k+1})}$$

on the interval (x_k, x_{k+1}) (Lutz Lehmann, 2019). It is readily seen to be continuous due to it being a composition of individually continuous functions. We now consider the extension of $\psi_k(x)$on $[x_k, x_{k+1}]$, namely,

$$\omega_k(x) = \lim_{x' \to x} \psi_k(x')$$

and I claim that it is continuous. Since $\psi_k(x)$is continuous on (x_k, x_{k+1}),

$$\omega_k(x) = \lim_{x' \to x} \psi(x') = \psi_k(x).$$

Now, on $\{x_k, x_{k+1}\}$,one considers

$$\omega_k(x) = \lim_{x' \to x} \psi(x')$$

$$= \lim_{x \to x'} \frac{f(x') - P_k(x')}{(x' - x_k)(x' - x_{k+1})}$$

$$= \lim_{x \to x'} \frac{f^{(1)}(x') - P_k^{(1)}(x')}{2x' - (x_k + x_{k+1})} \text{ (L'Hopital's Rule)}$$

$$= \frac{f^{(1)}(x) - P_k^{(1)}(x)}{2x - (x_k + x_{k+1})}.$$

One observes that both exist; since the numerator and the denominator evaluate to zero at $\{x_{k,} x_{k+1}\}$by virtue of $P_k(x)$being an interpolating polynomial, the use of L'Hopital's Rule is justified.

Thus, the function $\omega_k(x)$ takes the form

$$\omega_k(x) = \begin{cases} \dfrac{f(x) - P_k(x)}{\left(x - x_k\right)\left(x - x_{k+1}\right)}, & x \in \left(x_k, x_{k+1}\right) \\ \\ \dfrac{f^{(1)}\left(x_k\right) - P_k^{(1)}\left(x_k\right)}{\left(x_k - x_{k+1}\right)}, & x = x_k \\ \dfrac{f^{(1)}\left(x_{k+1}\right) - P_k^{(1)}\left(x_{k+1}\right)}{\left(x_{k+1} - x_k\right)}, & x = x_{k+1} \end{cases}$$

Thus, integrating on the given domain,

$$\int_{x_k}^{x_{k+1}} f(x)dx = \int_{x_k}^{x_{k+1}} P_k(x)dx + \int_{x_k}^{x_{k+1}} \omega_k(x)(x - x_k)(x - x_{k+1})\, dx.$$

From the above, we observe that $(x - x_k)(x - x_{k+1})$does not change sign in the interval and also keep in mind the fact that $\omega_k(x)$is continuous on $[x_k, x_{k+1}]$. This will enable us to use Theorem 2.1.

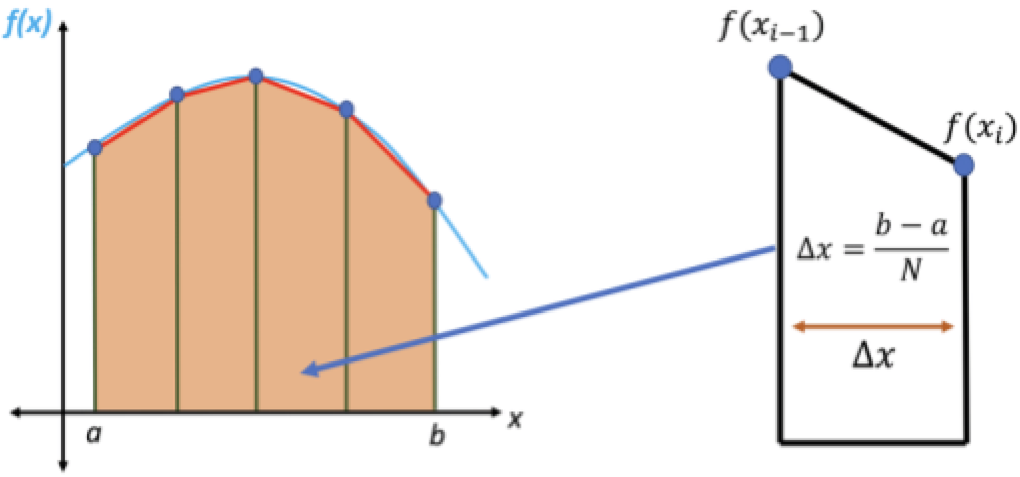


Figure 2: Trapezium rule with step size h = Δx, applied to a function f(x) with N = 4 times (ResearchGate, 2019).

Thus, by Theorem 2.1, for all k, there exists a $c_k \in (x_k, x_{k+1})$ such that

$$\int_{x_k}^{x_{k+1}} f(x)dx = \int_{x_k}^{x_{k+1}} P_k(x)dx + \omega_k(c_k) \int_{x_k}^{x_{k+1}} (x - x_k)(x - x_{k+1})\, dx.$$

Now, observing from Theorem 1.2 that there exists a $\xi_{c_k} \in (x_k, x_{k+1})$ such that

$$\omega_k(c_k) = \frac{f^{(2)}(\xi_{c_k})}{2!}.$$

Hence, we obtain that

$$\int_{x_k}^{x_{k+1}} f(x)dx$$

$$= \int_{x_k}^{x_{k+1}} P_k(x)dx + \frac{f^{(2)}(\xi_{c_k})}{2} \int_{x_k}^{x_{k+1}} (x - x_k)(x - x_{k+1})\, dx \text{ with } \xi_{c_k} \in (x_k, x_{k+1})$$

$$= \int_{x_k}^{x_{k+1}} P_k(x)dx - \frac{f^{(2)}(\xi_{c_k})}{2}\frac{(x_{k+1}-x_k)^3}{6}$$

$$= \int_{x_k}^{x_{k+1}} P_k(x)dx - \frac{f^{(2)}(\xi_{c_k})}{12}h^3,$$

after an elementary integration which is omitted (This is to be compared with the general result in Section 2.3).

Now, summing over all intervals $[x_k, x_{k+1}]$,

$$\int_a^b f(x)dx = \sum_{k=1}^{N} \left(\int_{x_k}^{x_{k+1}} P_k(x)dx - \frac{f^{(2)}(\xi_{c_k})}{12}h^3 \right)$$

$$= T(f,h) - \sum_{k=1}^{N} \frac{f^{(2)}(\xi_{c_k})}{12}(b-a)h^2$$

$$= T(f,h) + \sum_{k=1}^{N} \frac{f^{(2)}(\xi_{c_k})}{12}(a-b)h^2.$$

Since f has a continuous second derivative and

$$\min_{x\in[a,b]} f^{(2)}(x) \le \frac{1}{N}\sum_{k=1}^{N} f^{(2)}\left(\xi_{c_k}\right) \le \max_{x\in[a,b]} f^{(2)}(x),$$

by the Intermediate Value Theorem, there is a $\eta \in (a,b)$ such that

$$\sum_{k=1}^{N} f^{(2)}(\xi_{c_k}) = f^{(2)}(\eta).$$

Finally,

$$\int_a^b f(x)dx = T(f,h) + f^{(2)}(\eta)\frac{(a-b)h^2}{12}$$

$$= \frac{h}{2}\left(f(x_0) + 2\sum_{i=1}^{N-1} f(x_i) + f(x_N)\right) + f^{(2)}(\eta)\frac{(a-b)h^2}{12},$$

and we are done. ▨

(Numerical Integrations, 2018)

One can see from the above result, that the error depends on the square of the interval subdivision length, h. In practical terms, this means by doubling the number of points, one (roughly) reduces the error by a factor of four.

Before closing this section and moving onto the stochastic numerical methods, we can ascertain whether our approximation of the quadrature is an underestimate or an overestimate based on the concavity of the function on the interval. This is because the error depends on the second derivative of the function under consideration. So, assuming that $f$is a positive function, $a < b$ and that, $f^{(2)}(x) < 0$ for all x in a particular interval, then the error $\varepsilon = f^{(2)}(\eta)\frac{(a-b)h^2}{12} = -f^{(2)}(\eta)\frac{(b-a)h^2}{12} =- f^{(2)}(\eta) > 0$. This means that the approximation is an underestimate because of the positive correction one has to make to obtain the integral, that is the area under the curve. This confirms one's geometric intuitions; for instance, in figure 2, the function is perceived to be concave and clearly, the area under the trapezoids seems to underestimate the area under the curve.

2.3 Newton Cotes Formulas Error for the General Case

Now it is time to move on to the main result of this section. We will interpolate a function at $n + 1$ different points in the given interval of integration using a polynomial of degree n; then we will derive the formula for the quadrature and error by means of integration. From there, we will derive the quantities thereof for the composite Trapezium Rule by partitioning the interval in equally separated points and applying the quadrature and error formulas to each subinterval where the interpolating polynomials are linear, namely of degree one. This is done for the sake of concreteness.

We are given a function $f(x)$which is (n+1) times continuously differentiable on [a,b]; we fix points $\alpha \leq x_0 < x_1 < ... < x_n \leq \beta$ in $[\alpha, \beta]$where we will interpolate $f(x)$. Our goal is to evaluate (find an approximation with an error term) the expression

$$\int_{\alpha}^{\beta} f(x)dx.$$

By an argument similar to the one in Section 2.2, one can show that there exists a function ω(x) that is continuous on [α,β] (with a little care needed at the points of interpolation, resolved by taking limits) such that

$$f(x) - p_n(x) = \prod_{j=0}^{n} (x - x_j)\omega(x),$$

where $\omega(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}$, for some $\xi_x \in (\alpha, \beta)$ by Theorem 1.2.

Consequently,

$$\int_{\alpha}^{\beta} f(x)dx = \int_{\alpha}^{\beta} (p_n(x) + \prod_{j=0}^{n} (x - x_j)\omega(x))dx$$

$$= \int_{\alpha}^{\beta} p_n(x)dx + \int_{\alpha}^{\beta} \omega(x) \prod_{j=0}^{n} (x - x_j)dx$$

$$= \int_{\alpha}^{\beta} (\sum_{j=0}^{n} f(x_j)L_{n,j})dx + \int_{\alpha}^{\beta} \omega(x) \prod_{j=0}^{n} (x - x_j)dx$$

$$= \sum_{j=0}^{n} f(x_j) \int_{\alpha}^{\beta} \prod_{k=0,\, k\neq j}^{n} \frac{x-x_k}{x_j-x_k}dx + \int_{\alpha}^{\beta} \omega(x) \prod_{j=0}^{n} (x - x_j)dx$$ (by Theorem 1.1).

Let

$$l_j = \int_{\alpha}^{\beta} \prod_{k=0,\, k\neq j}^{n} \frac{x-x_k}{x_j-x_k}dx,$$

then the above can be written more compactly as

$$\int_{\alpha}^{\beta} f(x)dx = \sum_{j=0}^{n} f(x_j)\, l_j + \int_{\alpha}^{\beta} \omega(x) \prod_{j=0}^{n} (x - x_j)dx,$$

which is our desired result.
The quadrature rule is given by

$$\sum_{j=0}^{n} f(x_j)\, l_j$$

and the exact error is given by

$$\int_{\alpha}^{\beta} \omega(x) \prod_{j=0}^{n} (x - x_j) dx.$$

However, the above result for the error is not very versatile due to the intractable behaviour of the term

$$\prod_{j=0}^{n} (x - x_j),$$

which may or may not change signs in the interval [α,β]. One could however, try and split the integral into intervals where the product does not change sign and simplify things by assuming that the interpolation points are equally spaced out by a distance h, for instance. In any case, due to the continuity of ω(x) on a closed and bounded interval [α,β] one can bound it with a real number M (Extreme Value Theorem, Wolfram MathWorld). Then, one would obtain a more useful result for the error ε, namely,

$$|\varepsilon| = \left|\int_{\alpha}^{\beta} \omega(x) \prod_{j=0}^{n} \left(x - x_j\right) \mathrm{d}x\right| \leq \int_{\alpha}^{\beta} \left|\omega(x) \prod_{j=0}^{n} \left(x - x_j\right)\right| \mathrm{d}x$$
$$= \int_{\alpha}^{\beta} |\omega(x)| \left|\prod_{j=0}^{n} \left(x - x_j\right)\right| \mathrm{d}x \leq M \int_{\alpha}^{\beta} \left|\prod_{j=0}^{n} \left(x - x_j\right)\right| \mathrm{d}x.$$

Where the above comes from the following:

For any continuous function f on [a,b], one trivially has

$$-|f(x)| \leq f(x) \leq |f(x)|$$

which implies

$$-\int_{a}^{b} |f(x)| \mathrm{d}x \leq \int_{a}^{b} f(x) \mathrm{d}x \leq \int_{a}^{b} |f(x)| \mathrm{d}x,$$

whence

$$\left|\int_a^b f(x)\,\mathrm{d}x\right| \le \int_a^b |f(x)|\,\mathrm{d}x$$

However, the above warrants an analysis of the term

$$\prod_{j=0}^{n} (x - x_j).$$

It becomes clear that the above result is of purely theoretical interest as it is not very practical for large n.

3 Stochastic Methods of Numerical Integration

Having done a preliminary error analysis to justify the commonly used methods of deterministic numerical integration, in this case, Newton Cotes formulas - more specifically, the linear interpolant one, I deem that it is the right time to change perspective on the matter of numerical integration and introduce a stochastic element into it and obtain another numeric method for integration, namely the Monte Carlo method.

However, in the beginning of this transition, I started to wonder whether the above was feasible. I questioned how much further than the mathematics I was taught at school I had to go. I was lucky enough to only require the use of standard results in probability theory that are accessible to anyone with a strong interest in the subject. With that in mind, for a fruitful comprehension of the next section, a background in continuous random variables is beneficial; at any rate, the reader is invited to consult Lim, N. (2017). My goal is to arrive at a statistical ‘error’ expressed in the **language of probability**. This necessitates the construction of a random variable intuitively incorporating the notion of an integral which will be motivated below.

First, let us imagine like before, that we have a function $f(x)$, defined over a domain $[a, b]$. A particularly useful notion to have in mind is the fact that the integral of *f(x)* over [a,b] can be approximated using Riemann Sums, following Weisstein, Riemann Sum, of the form

$$\sum_{k=1}^{N} f(x_i^*)\Delta x_i$$

where the interval $[a, b]$is partitioned by N intervals of size Δx_i and the $x_i^*$are points in Δx_i.If the interval is split uniformly, then one has $\Delta x_i = \frac{(b-a)}{N}$and hence

$$\sum_{k=1}^{N} f(x_i^*)\Delta x_i = \frac{(b-a)}{N}\sum_{k=1}^{N} f(x_i^*)$$

My next step is introducing the stochastic element by defining for each interval Δx_i the continuous random variable X_i that is uniformly distributed on $[a, b]$. One may motivate this by thinking of the X_i as a 'random' pick of an x-value from the domain of integration [a,b]. I observe that all the X_i are identically distributed and are intended to be independent. We proceed and define the random variable $\langle F^N \rangle$ to be

$$\frac{(b-a)}{N}\sum_{i=1}^{N} f(X_i).$$

Intuitively, we begin by making N independent 'experiments' on the continuous random variables X_i, namely values of the x-axis. The estimate $\langle F^N \rangle$ is then the average over each rectangular approximation of the function evaluated at the outcome of our experiment times the length of the interval [a,b] (Scratchapixel) (see figure 3 for the case N=4).

This and the following sections are dedicated to showing that the random variable $\langle F^N \rangle$ does indeed correspond to an estimate of the integral of the function. This will be achieved by showing that its mean is the exact value of the integral

$$\int_a^b f(x)dx.$$

The above will enable us to prove some additional statements about the probability distribution of the absolute value of the difference between $\langle F^N \rangle$,our estimate and its mean *F,* the integral we want to calculate leading to a probabilistic expression for the 'error'.

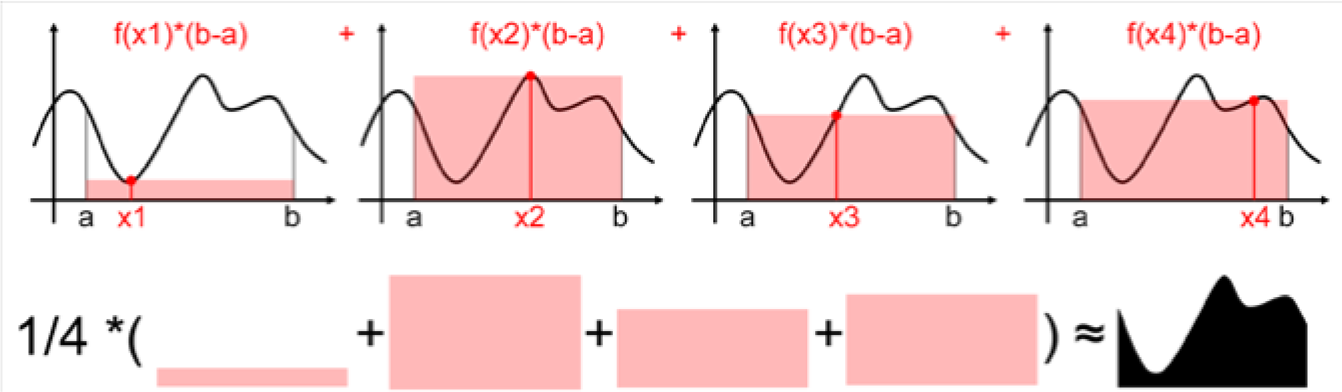


Figure 3: This is an example of how the integral could be estimated using a random process. In this case, N=4 (Scratchapixel, 2015).

3.1 Monte Carlo Integration

Lemma 3.1 *Let* $F = \int_a^b f(x)dx$*. Then, the expected value of* $\langle F^N \rangle$*,* $E[\langle F^N \rangle]$ equals *F.* (Scratchapixel, 2015)

Proof

Since the $\{X_i\}$ *are independent, we can exploit the linearity of the expectation (see* Lim, N. (2017), page 7*) and obtain,*

$$E[\langle F^N \rangle] = E[\frac{(b-a)}{N} \sum_{i=1}^{N} f(X_i)]$$

$$= \frac{(b-a)}{N} NE[f(X)]$$

$$= (b - a) \int_a^b f(x)pdf(x)dx,$$

but since the $\{X_i\}$are identically distributed, $pdf(x) = \frac{1}{(b-a)}$. This finishes the proof.

(Scratchapixel, 2015)

Lemma 3.2 *The variance of* $\langle F^N \rangle$, $Var[\langle F^N \rangle] = \frac{(b-a)^2}{N} Var(X_i)$

Proof

Since $\{X_i\}$ are independent, $Var[\langle F^N \rangle]$

$$= Var[\frac{(b-a)}{N} \sum_{i=1}^{N} f(X_i)]$$

$$= \frac{(b-a)^2 N}{N^2} Var[f(X_i)],$$

by basic properties of the variance which finishes the proof. ▨

Now, we would like to prove a statement about the 'spread' of the distribution of the errors given a sample size N. This will be done once we will establish the following important result.

3.2 Chebyshev's Inequality

Lemma 3.3 ***(Chebyshev's Inequality)*** *Let X a continuous random variable with probability density function* $f(x)$, $\varepsilon > 0$, *with* $E[x] = \mu$ *and* $\sigma = \sqrt{Var[X]}$, *then the probability* $Pr(|X - E[X]| \geq \varepsilon\sigma) \leq \frac{1}{\varepsilon^2}$.

(This is illustrated in figure 4.)

(Zhengtianyu, 2014)

Proof

From the definition of variance,

$$\sigma^2 = \int_a^b (x-\mu)^2 f(x)\,\mathrm{d}x$$

$$= \int_{[a,b]\cap\, |X-\mu| \geq \sigma\varepsilon} (x-\mu)^2 f(x)\,\mathrm{d}x$$

$$+ \int_{[a,b]\cap\, |X-\mu| < \sigma\varepsilon} (x-\mu)^2 f(x)\,\mathrm{d}x$$

$$\geq \int_{[a,b]\cap\, |X-\mu| \geq \sigma\varepsilon} (x-\mu)^2 f(x)\,\mathrm{d}x$$

because the second term is the integral of a positive function and hence the Riemann integral is itself positive - it is interpreted as a probability which by definition, is positive.

If $x$is part of the domain of the set $|X - \mu| \geq \sigma\varepsilon$, then the previous condition implies quite readily that $(x - \mu)^2 \geq \sigma^2\varepsilon^2$, upon squaring both sides.

Thus,

$$\sigma^2 \geq \int_{[a,b]\cap\, |X-\mu| \geq \sigma\varepsilon} (x-\mu)^2 f(x)\,\mathrm{d}x$$

$$\geq \int_{[a,b]\cap\, |X-\mu| \geq \sigma\varepsilon} \sigma^2\varepsilon^2 f(x)\,\mathrm{d}x$$

$$= \sigma^2\varepsilon^2 \int_{[a,b]\cap\, |X-\mu| \geq \sigma\varepsilon} f(x)\,\mathrm{d}x$$

$$= \sigma^2\varepsilon^2 \mathbb{P}\mathrm{r}(|X-\mu| \geq \sigma\varepsilon)$$

This finishes the proof. ▨

(Zhengtianyu, 2014)

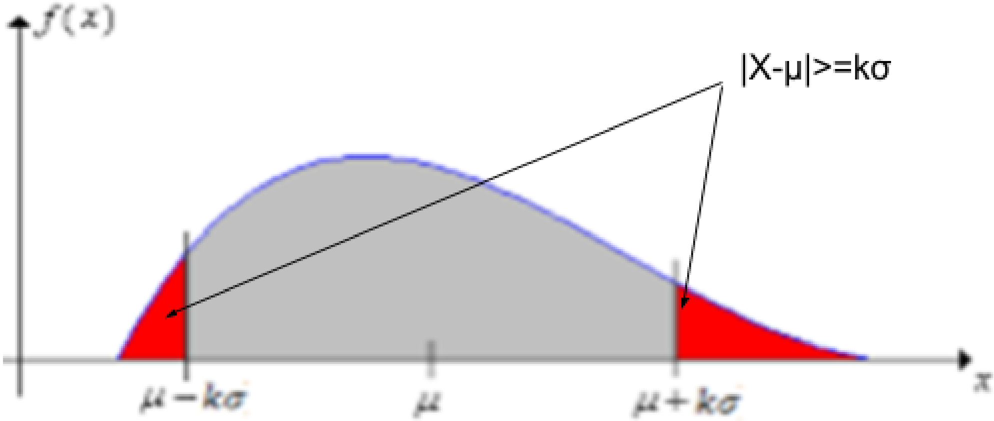


Figure 4: A hypothetical probability distribution where the real constant k corresponds to our ε. The area in red displays the actual probability (Lemma 3.3), whereas lemma 3.3 gives an upper bound, namely, Chebyshev's Inequality. (Ques10, 2016)

3.3 Probabilistic Error for Monte Carlo Integration

We now make the substitution $\delta = \frac{1}{\varepsilon}$ and Lemma 3.3 becomes $Pr(|X - \mathrm{E}[X]| \geq \frac{\sigma}{\delta}) \leq \delta^{2}$.

From Lemma 3.2, $\sigma = \frac{(b-a)}{\sqrt{N}}\sqrt{Var[f(X)]}$, thus, for arbitrary $\delta > 0$, we have

$$\mathbb{Pr}\left(\left|\langle F^N \rangle - F\right| \geqslant \frac{(b-a)\sqrt{Var[f(X)]}}{\sqrt{N}\delta}\right) \leqslant \delta^2$$

This is a major result as we introduced sample size dependence and statistically bounded our error. This can be interpreted as follows: we can make as small as we wish the probability that our estimation differs from its mean (the exact value of the integral) by an amount (which we can make as small as we wish) provided we pick *N* big enough. In practical terms, considering Lemma 3.2, “four times more samples are needed to reduce the error of the estimate by half” (Scratchapixel, 2015).

4 A representative example of comparison of errors resulting from both approaches

With all the theoretical results obtained, a return to the real world is warranted in order to better explore the errors furnished by the Deterministic Newton Cotes and the Stochastic Monte Carlo Methods (Section 3.3). Concerning the former, I will be using the composite Trapezium Rule (Section 2.2). Thus, a representative example will be used and a comparison made. To be more precise, the comparison will comprise of observing the behaviour of the error of the integral with an increase in the number of 'steps' or computations made, N where both expressions for the error in Sections 2.3 and 3 have a dependence on it. Not to mention, in determining the order of the error, I shall proceed and use the big-O notation following the definition given by (Big-O, NIST)[1]. The function that will be considered is

$$f(x) = e^{-x^2}$$

in the interval $[-2, 2]$(figure 5). The goal is to evaluate numerically and compare the errors yielded by the above methods, of the integral I,which equals

$$I = \int_{-2}^{2} f(x)dx = \int_{-2}^{2} e^{-x^2} dx .$$

I was led to consider the function $f(x)$because of its representative nature as it is frequently encountered in probability and physics where many statistical and physical systems contain variables that exhibit the behaviour of $f(x)$. Also, the above function does not have an elementary antiderivative and hence its integral cannot be evaluated analytically. A prominent and significant example is its use to model the Bit Error Rate, that is measuring the degree of integrity when a message is transmitted in a communication system (Error Functions and Bit Error Rate). In addition, it can also be used to approximate the value of π, as by graphical means (see figure 5), one can proceed and approximate the integral

$$\int_{-\infty}^{\infty} e^{-x^2} dx = 2\sqrt{\pi}$$

Cirrito, F. (2009), page 561,

with

$$\int_{-2}^{2} e^{-x^2} dx.$$

[1] A function f(x) is said to be of order O(g(x)) denoted f(x)~O(g(x)) if there are positive constants c and k such that 0≤f(x)≤cg(x) for all x≥k

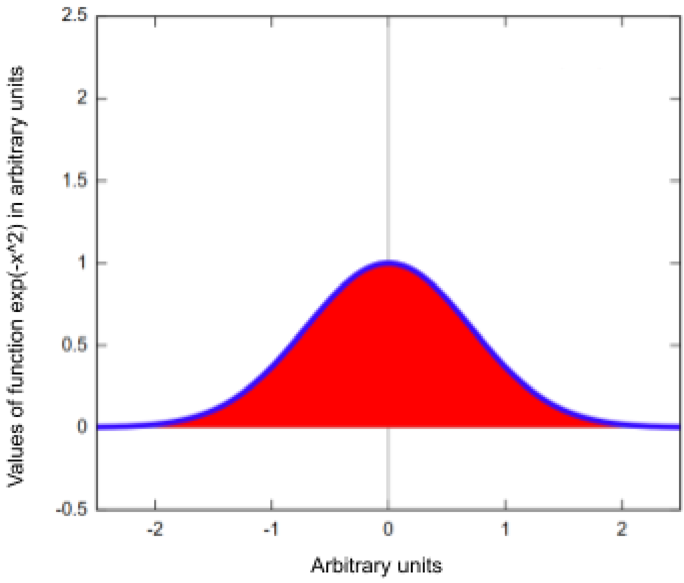


Figure 5: the red area represents the the integral I.
The blue curve represents the function under consideration, exp(-x^2).
(https://upload.wikimedia.org/wikipedia/commons/2/2f/E%5E%28-x%5E2%29.svg)

4.1 Using the Composite Trapezoid rule

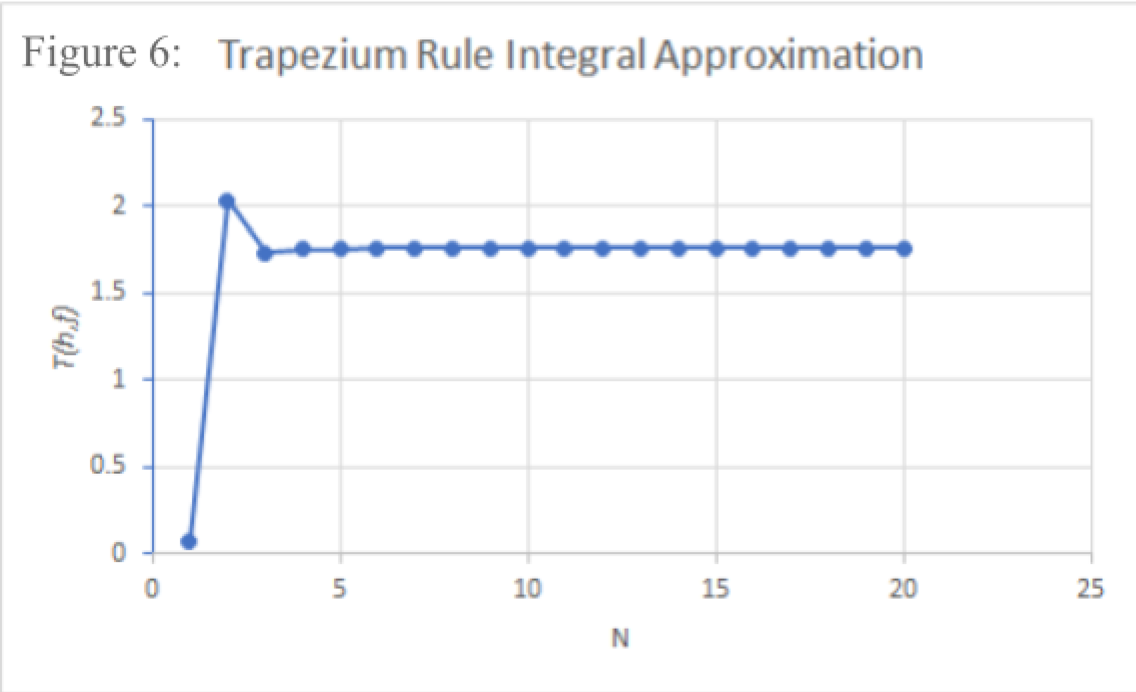
In any case, let us begin the evaluation of the integral using the deterministic quadrature rule, namely, the Composite Trapezoid Rule. From Section 2.2, we let

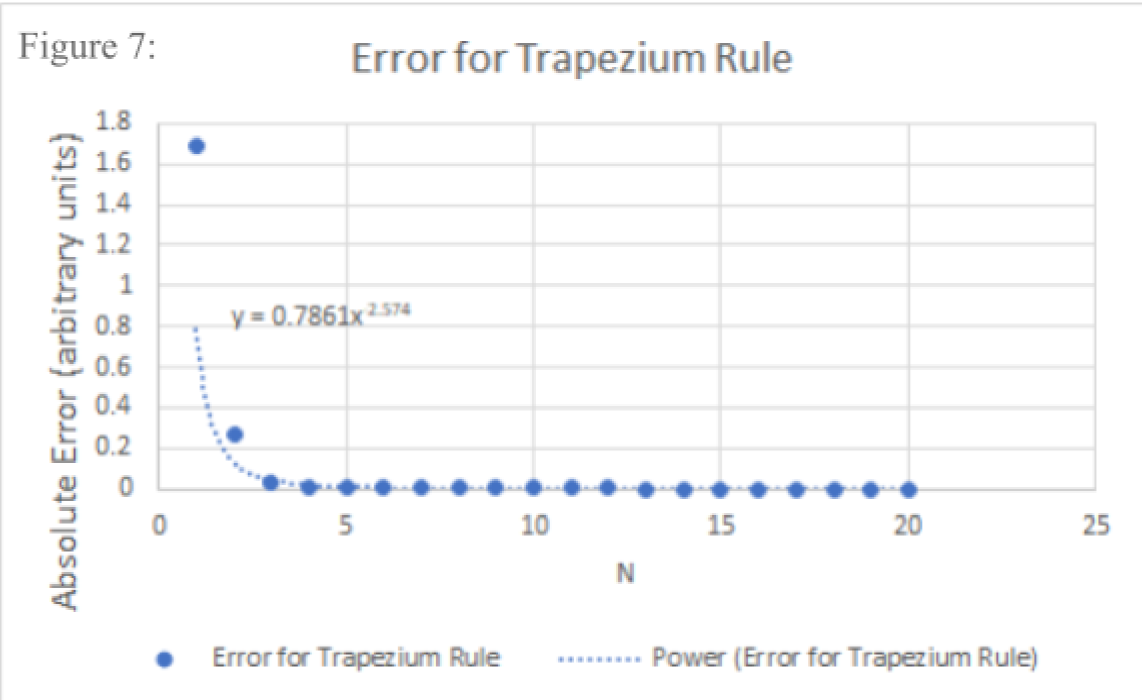
$$a = -2,\ b = 2,\ h = \frac{b-a}{N},\ and\ f = e^{-x^2}$$

where N corresponds to the **number of trapezia.** I then proceed to evaluate the integral using the approximation

$$T(h, f) = T(\frac{2-(-2)}{N}, e^{-x^2}) = T(\frac{4}{N}, e^{-x^2})$$

by incrementing N from 1 to 20.





I used the computational knowledge engine *WolframAlpha*[2] to obtain these values and the result is displayed in figure 6 (Table 1 in Appendix). It is readily seen that $T(h, f)$, the approximation to the integral reaches a

[2]Trapezoidal Rule Calculator. (n.d.).

constant value, namely, that of the integral I which is approximately 1.7642[3]. However, we need an expression for the absolute error. This is simply achieved by plotting the absolute value of $T(h, f)$minus the above value for the integral I (figure 7 and Table 1 in Appendix).

I shall now demonstrate that the absolute error is of order $\frac{1}{n^2}$. Using the definition in footnote 1 by (Big-O, NIST), I shall demonstrate that the best fit power trend line of the graph for the absolute error (figure 7) is bounded by some constant for all values of x greater than a real constant. Thus, I let $f(x) = 0.7861x^{-2.574}$, $g(x) = \frac{1}{x^2} = x^{-2}$ and proceed by letting c=1 and k = 1. Now, verifying the condition in the footnote, I obtain

$$0 \leq 0.7861x^{-2.574} \leq x^{-2.574} \leq x^{-2},\ \forall x \geq 1.$$

Thus, I deduce that $Error_{Trapezium}(x) \sim O(g(x))$or by substituting N for x, one has $Error_{Trapezium}(N) \sim O(\frac{1}{N^2})$in agreement with the previous theory in Section 2.2.

4.2 Using the Stochastic Monte Carlo Method

Similarly for the stochastic case, from Section 3, we will proceed to approximate I by plotting values of the random variable

$$< F^N > = \frac{(b-a)}{N} \sum_{i=1}^{N} f(X_i),$$

where $a = -2$, $b = 2$and N represents the number of values taken from the domain of the function $f(x) = e^{-x^2}$. I thus obtain

$$< F^N > = \frac{(2-(-2))}{N} \sum_{i=1}^{N} e^{-X_i^2} = \frac{4}{N} \sum_{i=1}^{N} e^{-X_i^2}$$

and

$$F = \int_{-2}^{2} e^{-x^2}\, dx$$

where the $X_i$are independent and identically distributed uniform continuous random variables on the interval $[a, b] = [-2, 2]$. I will plot using the Microsoft Excel software, a value of $< F^N >$upon random generation of the $X_i$using the RAND function[4] for each N from 1 to 20 (figure 8 and Table 2 in Appendix) and then separately from 1 to 100 - incrementing N by 5 each time (figure 9- the orange line indicates the 'exact' value of the integral and Table 3 in Appendix). Why the need for two ranges, one may ask? Well, I decided to plot randomly generated values of $< F^N >$ for increasing N on two separate scales, because it became apparent that

[3] Integral Calculator.
[4] RAND function.

the convergence rate of the above method is slower compared to that of the deterministic method (Trapezium Rule). For the small N case (figure 10 and Table 2 in Appendix), there appears to be non real pattern except for an erratic and gradual decrease in the absolute error. So, in order to visualise the statistical algorithm, I had to probe larger values of N as aforementioned.

Figure 8: Stochastic method Integral for Small N

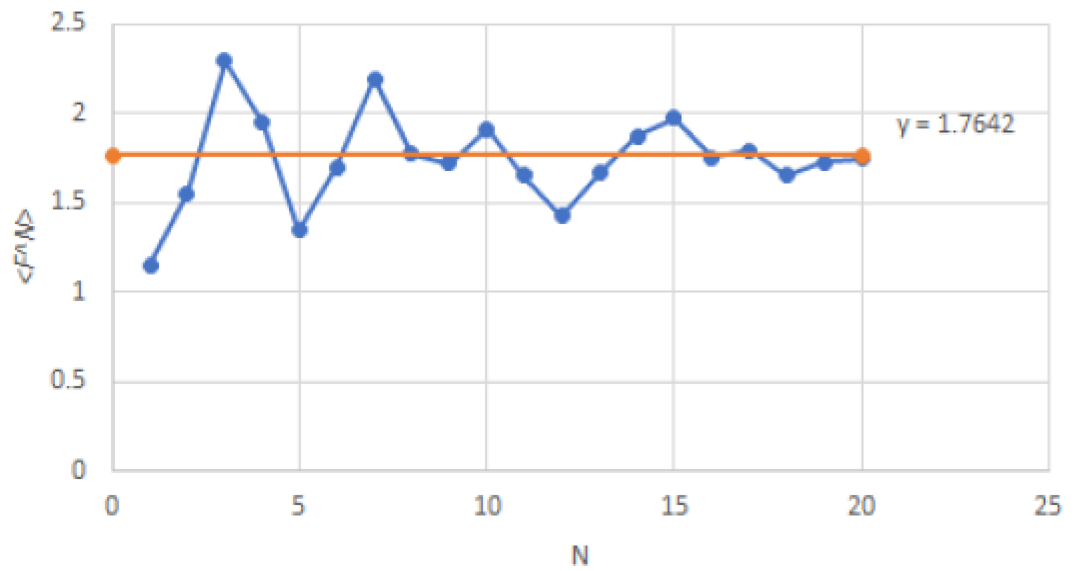


Figure 9: Stochastic method Integral for Large N

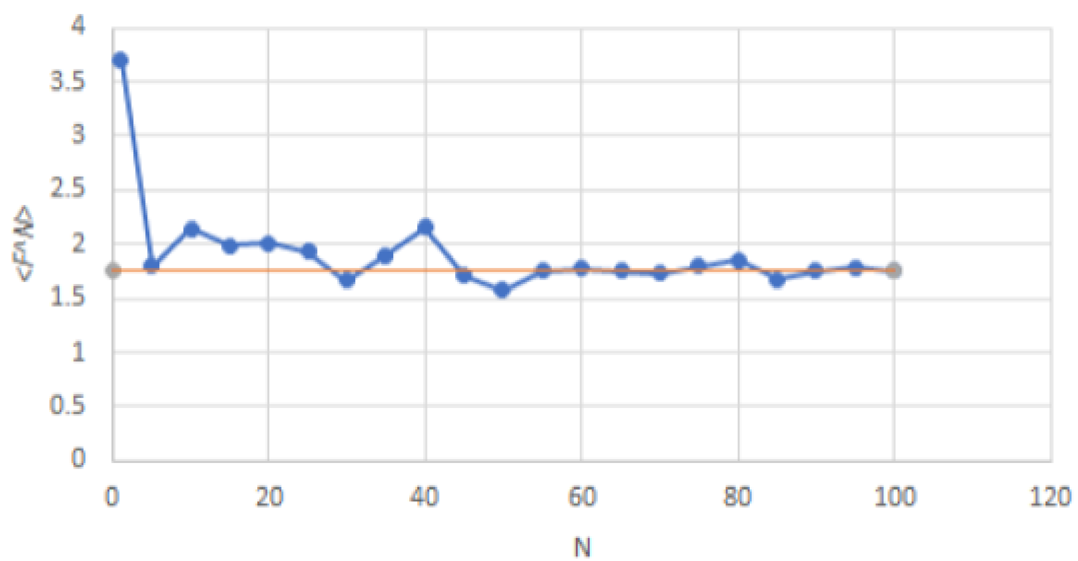


Figure 10: Stochastic method Error for Small N

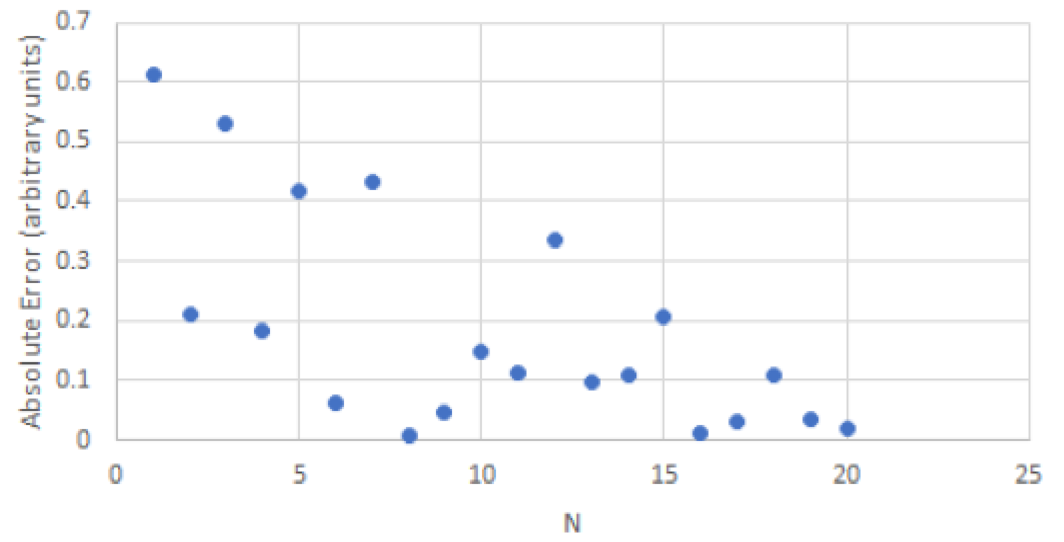
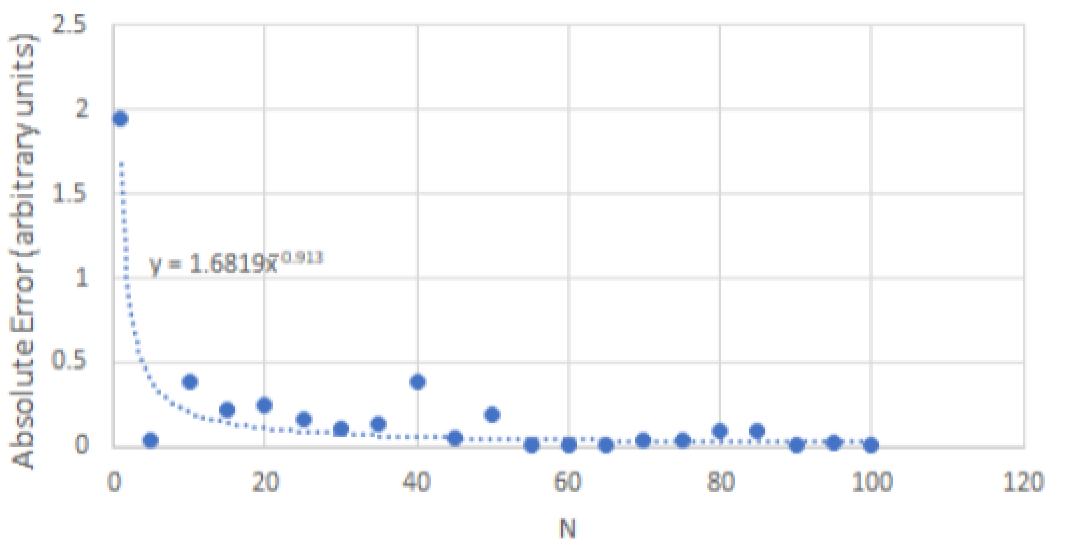


Figure 11: Error series for Large N



I shall now demonstrate that the absolute error for the stochastic method is of order $\frac{1}{\sqrt{N}}$. Using the definition in footnote 1 by (Big-O, NIST), I shall demonstrate that the best fit power trend line of the graph for the absolute error (figure 11 and Table 3 in Appendix) is bounded by some constant for all values of x greater than a real constant. Thus, I let $f(x) = 1.6819x^{-0.913}$, $g(x) = \frac{1}{\sqrt{x}} = x^{-0.5}$ and proceed by letting c=2 and k = 1. Now, verifying the condition in the footnote, I obtain

$$0 \leq 1.6819x^{-0.913} \leq 2x^{-0.913} \leq 2x^{-0.5},\ \forall x \geq 1.$$

Thus, I can say that $Error_{Stochastic}(x) \sim O(g(x))$or by substituting N for x, one has $Error_{Stochastic}(N) \sim O(\frac{1}{\sqrt{N}})$in agreement with the theory in Section 3.3.

5 Conclusion

What have I learnt from the above deliberations? Well, my central finding with regards to the deterministic errors for numerical integration in Section 2 is

$$\int_{\alpha}^{\beta} \omega(x) \prod_{j=0}^{n} (x - x_j) dx$$

where

$$\omega(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}$$

for some $\xi_x \in (a, b)$ for $x \in (x_k, x_{k+1})$(see Section 2.2).

For the Monte Carlo method, the probabilistic error estimate I deduced is given by

$$\mathbb{Pr}\left(\left| \langle F^N \rangle - F \right| \geqslant \frac{(b-a)\sqrt{Var[f(X)]}}{\sqrt{N}\delta} \right) \leqslant \delta^2$$

(see Section 3.3).

The result for the deterministic case has an advantage in its generality as the result if for an arbitrary degree, n. However, it is not very flexible; this is because each individual case of interpolating polynomial with different degrees requires a different treatment due to the implicit dependence of ξ on x. This entails that an exact determination of the error of numerical integration is impossible. This is because of the nature of the variable ξ, whose nature is completely obscure; this is due to the nature of the proof of Theorem 1.2, which is a **non-constructive proof**. That is, one does not provide an explicit algorithm for obtaining the result, in this case ξ; instead, one proves its existence (Eric W.). It is noteworthy to comment that some mathematicians find this proof method to be objectionable. This may raise some knowledge issues as it may have implications for the validity or the quality of the mathematical truth thereof. A possible solution, albeit depending on function f(x) itself, would be to **place bounds** on $f^{(n+1)}(\xi_x)$ - where n is the number of interpolation points - on the interval [α, β], as per Corollary 1.1, where one loses any dependence of ξ, at the **cost of exactitude**.

I recognise the above realisation as a limitation of my methodology and hope to have shared some light with the relatively simple example of the composite Trapezium Rule. Another limitation of this approach is that for large n, the error term becomes volatile (see figure 1) and a both natural and pertinent question is how to choose interpolation nodes that will minimise the error in some way. A possible extension leading to further research could include a discussion of **Chebychev nodes** and how they minimise the term

$$\prod_{j=0}^{n} (x - x_j)$$

in the error (Ron, A., 2010). Alternatively, one could have chosen an entirely different way of performing numerical integration and have decided to use the method of Gaussian Integration. One would proceed similarly to the Newton Cotes method of finding a polynomial, namely, a Legendre polynomial, to approximate the function of interest, albeit in the Gaussian case, it is chosen optimally (Dyer, C., 2002).

Now, it is time to evaluate the results of Section 4 by comparing the Composite Trapezoid Rule to the Statistical Monte Carlo method from Section 3. From Section 3, I inferred that the dependence on the sample size for the latter is of the order $\frac{1}{\sqrt{N}}$ through the inclusion of such a factor in the probabilistic 'error term' given

$$\mathbb{P}\left(\left|\frac{4}{N}\left[\sum_{i=1}^{N} e^{-X^2_i}\right] - \int_{-2}^{2} e^{-x^2}\mathrm{d}x\right| \geqslant \frac{4\sqrt{Var\left(e^{-X_j^2}\right)}}{\sqrt{N}\delta}\right) \leqslant \delta^2$$

by

upon making relevant substitutions (see Sections 3.3 and 4.2) where j is an arbitrary integer from 1 to N as the X_i are identically distributed, with the choice of i being immaterial. On the other hand, the error term for the Composite Trapezium Rule is given by

$$-f^{(2)}(\eta)\frac{(2-(-2))h^2}{12} = -f^{(2)}(\eta)\frac{4h^2}{12}$$

for some $\eta \in (-2, 2)$ (see Sections 2.2 and 4.1) and I deduce that the error is of order $\frac{1}{N^2}$ as it is dependent on the square of the interval length, h given by $\frac{(2-(-2))}{N} = \frac{4}{N}$. Thus, in theory, we get a '**faster' convergence for the deterministic method**. The above was verified in Sections 4.2 and 4.1 respectively. Restating my findings from Section 4, I obtained

$$Error_{Stochastic}(N) \sim O(\frac{1}{\sqrt{N}})$$

for the stochastic method and

$$Error_{Trapezium}(N) \sim O(\frac{1}{N^2})$$

for the Composite Trapezium Rule.

However, a closer yet more general investigation is warranted as the Monte Carlo method is much more easily generalisable to higher dimensional domains and retains the same order of statistical error, whereas the former deterministic methods have increasingly slower convergence rates (Scratchapixel, 2015). Not to mention, the 'cost' of the computation increases exponentially for deterministic quadrature rules; this is known as the *Curse of Dimensionality* and is a widespread phenomenon in computational mathematics (Novak, E., 1997).

Moreover, there is room for optimisation of the stochastic method warranting further research. Another extension of the investigation could include optimising the statistical method by a choice of another probability density function (pdf, see Section 3.1) for the random variable X(see proof of Lemma 3.1), this is known as *Variance Reduction* (Scratchapixel, 2015). Changing the probability distribution of the random variable X would entail changing the frequencies of selecting different values of X in the interval of interest. This could reduce computational costs and have the potential of optimising the above algorithm. However, the above Monte Carlo method depends on the ability to generate random numbers, a non trivial task (Scratchapixel) where biases of the (pseudo-) random number generation algorithm used in the Microsoft Excel program may have affected the convergence rate of the integral estimate (see figures 10,11).

6 Bibliography

Battista, N. (2019). *Illustration of the composite Trapezoid Rule*. photograph. Retrieved from https://www.researchgate.net/figure/Image-illustrating-the-the-composite-trapezoid-rule_fig2_331257075

Big-O, NIST. Retrieved from https://xlinux.nist.gov/dads/HTML/bigOnotation.html

Buckle, N., Dunbar, I., & Cirrito, F. (2009). *Mathematics higher level (core): International Baccalaureate*. Victoria: IBID Press.

Calculus. (2019, September 21). Retrieved from https://en.wikipedia.org/wiki/Calculus.

Dunham, W. (2018). *The calculus gallery: Masterpieces from Newton to Lebesgue* (Vol. 60). Princeton University Press.

Dyer, C. (2002, April 24). Gaussian Quadrature. Retrieved February 10, 2020, from http://pathfinder.scar.utoronto.ca/~dyer/csca57/book_P/node44.html

Error Functions and Bit Error Rate. (2008, August 21). Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470754443.app2

Explain Chebyshev's Inequality with suitable example. (n.d.). Retrieved from https://www.ques10.com/p/17129/explain-chebyshevs-inequality-with-suitable-exampl/.

Integral Calculator. Retrieved from https://www.integral-calculator.com/

Intermediate Value Theorem. (n.d.). Retrieved from http://mathworld.wolfram.com/IntermediateValueTheorem.html.

Jia, Y.-B. (2017, September 21). Polynomial Interpolation. Retrieved 3, 2019, from http://web.cs.iastate.edu/~cs577/handouts/interpolate.pdf.

Lambers, J. (2010). Lagrange Interpolation. Retrieved 3, 2019, from https://www.math.usm.edu/lambers/mat772/fall10/lecture5.pdf.

Lim, N. (2017). Continuous Random Variables. Retrieved August 3, 2019, from https://www.stat.washington.edu/~nehemyl/files/UW_MATH-STAT395_continuous-rrv.pdf.

Liouville's theorem (differential algebra). (2019, May 13). Retrieved from https://en.wikipedia.org/wiki/Liouville's_theorem_(differential_algebra).

Lutz Lehmann (https://math.stackexchange.com/users/115115/lutz-lehmann), On the Lagrange Interpolation Error for the Trapezoidal Quadrature Rule, URL (version: 2019-08-04): https://math.stackexchange.com/q/3158596

Novak, E., & Ritter, K. (1997, January 1). The Curse of Dimension and a Universal Method For Numerical Integration. Retrieved February 10, 2020, from https://link.springer.com/chapter/10.1007/978-3-0348-8871-4_15

Numerical Integrations. (2018, April 26). Retrieved August 4, 2019, from http://faculty.uml.edu/hung_phan/362numer/362_lec_15.pdf.

Overfitting. (2019, August 27). Retrieved from https://en.wikipedia.org/wiki/Overfitting.

RAND function. (n.d.). Retrieved from https://support.office.com/en-us/article/rand-function-4cbfa695-8869-4788-8d90-021ea9f5be73

Renze, John and Weisstein, Eric W. "Extreme Value Theorem." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/ExtremeValueTheorem.html

Ron, A., Cowlishaw, M., & Fillmore, N. (2010, October 12). Introduction to Numerical Analysis, Lecture 10: Introduction to Splines. Retrieved February 10, 2020, from http://pages.cs.wisc.edu/~amos/412/lecture-notes/lecture10.pdf

Scratchapixel. (2015, April 18). Retrieved from https://www.scratchapixel.com/lessons/mathematics-physics-for-computer-graphics/monte-carlo-methods-in-practice/monte-carlo-integration.

SpacedSpaced 1, & AlexanderAlexander 3. (1965, March 1). Looking for a way to improve my inductive proof of a statement derived by Rolle's Theorem. Retrieved from https://math.stackexchange.com/questions/1035627/looking-for-a-way-to-improve-my-inductive-proof-of-a-statement-derived-by-rolle.

Trapezoidal Rule Calculator. (n.d.). Retrieved from https://www.wolframalpha.com/widgets/view.jsp?id=40563c7ee6ef51c162833327f8a3880d

Trench, W. F. (2013, December). INTRODUCTION TO REAL ANALYSIS. Retrieved 4, 2019, from http://ramanujan.math.trinity.edu/wtrench/texts/TRENCH_REAL_ANALYSIS.PDF.

Weisstein, Eric W. "Riemann Sum." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/RiemannSum.html

Weisstein, E. (2019). Rolle's Theorem. Retrieved August 5, 2019, from http://mathworld.wolfram.com/RollesTheorem.html.

Weisstein, Eric W. "Nonconstructive Proof." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/NonconstructiveProof.html

Zhengtianyu. (2014, January 5). Proof of Chebyshev's Inequality. Retrieved from https://zhengtianyu.wordpress.com/2014/01/04/proof-of-chebyshevs-inequality/.

7 Appendix

Below are the data used for the figures in Section 4.

N	Value of Integral	Absolute Error
1	0.07326	1.690902782
2	2.03663	0.272467218
3	1.73424	0.029922782
4	1.75407	0.010092782
5	1.757	0.007162782
6	1.75914	0.005022782
7	1.76039	0.003772782
8	1.76124	0.002922782
9	1.76183	0.002332782
10	1.76226	0.001902782
11	1.76258	0.001582782
12	1.76283	0.001332782
13	1.76303	0.001132782
14	1.76318	0.000982782
15	1.7633	0.000862782
16	1.76341	0.000752782
17	1.76349	0.000672782
18	1.76356	0.000602782
19	1.76363	0.000532782
20	1.76368	0.000482782

Table 1: Data for figures 6 (columns 1 and 2) and 7 (columns 1 and 3).

N	Approximate Integral	Absolute Error
1	1.151	0.613162782
2	1.555	0.209162782
3	2.2935	0.529337218
4	1.946	0.181837218
5	1.3455	0.418662782
6	1.7011	0.063062782
7	2.196	0.431837218
8	1.7723	0.008137218
9	1.719	0.045162782
10	1.9109	0.146737218
11	1.6506	0.113562782
12	1.42777	0.336392782
13	1.666	0.098162782
14	1.8739	0.109737218
15	1.9701	0.205937218
16	1.754	0.010162782
17	1.796	0.031837218
18	1.6555	0.108662782
19	1.729	0.035162782
20	1.7438	0.020362782

Table 2: Data for figures 8 (columns 1 and 2) and 10 (columns 1 and 3).

N	Approximate Integral	Absolute Error
1	3.711	1.946837218
5	1.7988	0.034637218
10	2.148	0.383837218
15	1.987	0.222837218
20	2.014	0.249837218
25	1.932	0.167837218
30	1.6619	0.102262782
35	1.8993	0.135137218
40	2.155	0.390837218
45	1.7098	0.054362782
50	1.5719	0.192262782
55	1.751	0.013162782
60	1.7756	0.011437218
65	1.75786	0.006302782
70	1.7306	0.033562782
75	1.798	0.033837218
80	1.855	0.090837218
85	1.677	0.087162782
90	1.756	0.008162782
95	1.791	0.026837218
100	1.758	0.006162782

Table 3: Data for figures 9 (columns 1 and 2) and 11 (columns 1 and 3)