Newton's Discovery of

The General Binomial Theorem

by interpolation over integral sequences

V. Belde and V. Elgersma-Azmanov

Originally Submitted 10 March, 2026

Last Revision 06 August, 2026

Contents

1 Introduction

This report covers several of Isaac Newton's key writings pertaining to the discovery of the binomial theorem and its application to root extractions and finding areas under curves. In Section 2, we present our commentary on a page from Newton's handwritten private notebook[1], in which he appears to work out the power series for $\int \frac{1}{1+x} dx$ and $\int \sqrt{1-x^2} dx$ by finding a pattern in the series expansions of the more general set of integrals $\int (1+x)^n dx$ and $\int (1-x^2)^n dx$ for $n = 0, 1, 2, 3....$ In the first case, Newton *extrapolates* the pattern to the case $n = -1$. In the second case, he *interpolates* the pattern to the half-odd integer cases $n = \frac{1}{2}, \frac{3}{2}...$, by first finding a general expression for the coefficients in the pattern, in terms of integers n and m, and then substituting half-odd integer values for n. For our commentary, we have relied on scans hosted by the *Cambridge Digital Library*[2], as well as the transcription provided by Jacqueline Stedall[3].

In Section 3, we comment on a section on algorithmic root extraction from Newton's *De Analysi* (1711, pp. 6-7), for which we also rely on Stedall's transcription[4]. The long-hand

[1]Isaac Newton, 'Mathematical Papers (Cambridge University Library, MS Add. 3958:72)', England, 1665, https://cudl.lib.cam.ac.uk/view/MS-ADD-03958/139.

[2]Newton, 'Mathematical Papers (Cambridge University Library, MS Add. 3958:72)'.

[3]Jacqueline Stedall, *Mathematics Emerging: A Sourcebook 1540–1900* (Oxford University Press, 2008), p. 192, https://nx89456.your-storageshare.de/s/ArjNPeHYjkom6Zt.

[4]Stedall, *Mathematics Emerging: A Sourcebook 1540–1900*, p. 195.

extraction of roots algorithm was one of the procedures "much shortened" by Newton's binomial theorem[5]. In Section 4 and Section 5, we comment on two of the most famous letters in the history of science – the so-called *Epistola Prior*[6] and *Posterior*[7], in which Newton shares the Binomial Theorem with his contemporary Gottfried Wilhelm Leibniz (1646–1716) through the intermediary Henry Oldenburg.

2 Integration by interpolation and intercalation of series

> "In the winter between the years 1664 & 1665 upon reading Dr Wallis's Arithmetica Infinitorum & trying to interpole his progressions for squaring the circle, I found out first an infinite series for squaring the circle & then another infinite series for squaring the Hyperbola & soon after."
>
> — Newton, quoted by Whiteside[8]

[5]Isaac Newton, 'Epistola Prior (Letter to Gottfried Wilhelm Leibniz via Henry Oldenburg, June 13, 1676)', in *The Correspondence of Isaac Newton*, 2 (1676–1687), *The Correspondence of Isaac Newton*, ed. Herbert W. Turnbull (Cambridge University Press, 1960).

[6]Newton, 'Epistola Prior (Letter to Gottfried Wilhelm Leibniz via Henry Oldenburg, June 13, 1676)'.

[7]Isaac Newton, 'Epistola Posterior (Letter to Gottfried Wilhelm Leibniz via Henry Oldenburg, October 24, 1676)', in *The Correspondence of Isaac Newton*, 2 (1676–1687), *The Correspondence of Isaac Newton*, ed. Herbert W. Turnbull (Cambridge University Press, 1960).

[8]D. T. Whiteside, 'Newton's Discovery of the General Binomial Theorem', *The Mathematical Gazette* 45, no. 353 (1961): 175–80, https://doi.org/10.2307/3612767.

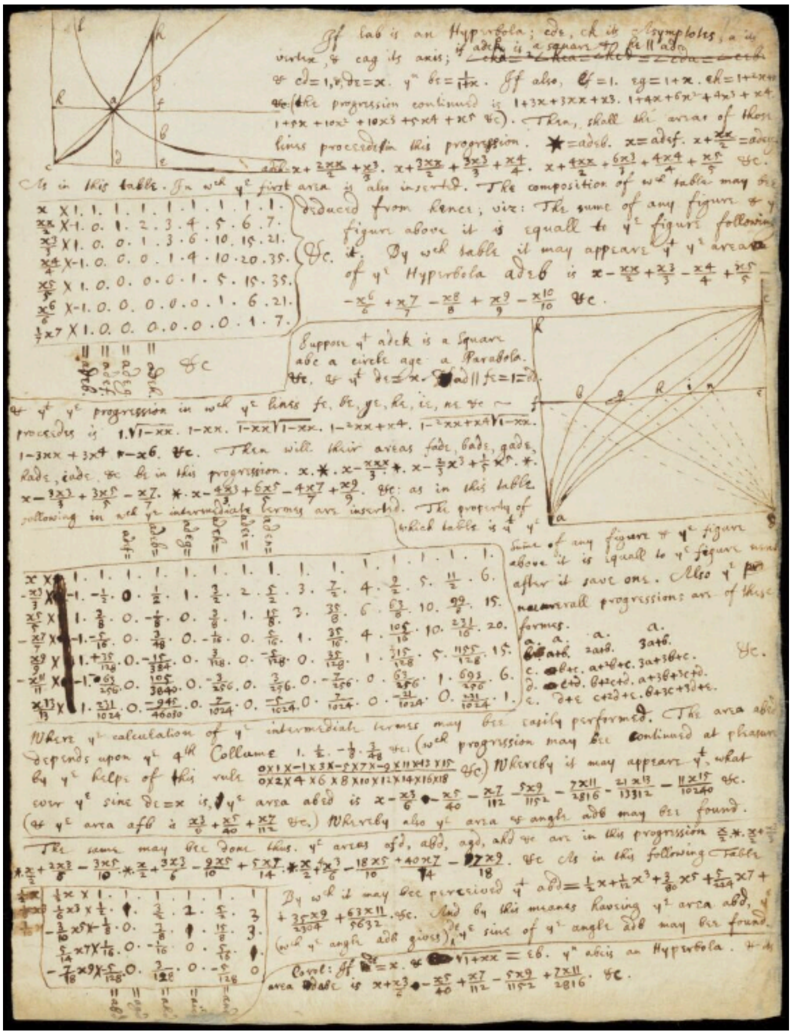


Figure 1: Newton, Isaac. 'Mathematical Papers (Cambridge University Library, MS Add. 3958:72)'. England, 1665. https://cudl.lib.cam.ac.uk/view/MS-ADD-03958/139

In this page from a young Newton's handwritten notes (Figure 1), we find the power series expansions of $\int \frac{1}{1+x}dx$ and $\int \sqrt{1-x^2}dx$, which is obtained by extrapolating the pattern created by the integrals $\int (1+x)^n dx$ and $\int (1-x^2)^n dx$, for positive integer n. We begin by looking at the integral of $\frac{1}{1+x}$, which he does in the top part of the manuscript.

2.1 Integral of the Hyperbola ($\int \frac{1}{1+x}dx$)

Newton begins by considering the following curves, sketching the first four (see Figure 2):

$$
\begin{aligned}
y &= \frac{1}{1+x} \\
y &= 1 \\
y &= (1+x) \\
y &= (1+x)^2 = 1 + 2x + x^2 \\
y &= (1+x)^3 = 1 + 3x + 3x^2 + x^3 \\
y &= (1+x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4 \\
y &= (1+x)^5 = 1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5
\end{aligned}
\tag{1}
$$

> If *lab* is an Hyperbola; *cde*, *ck* its Asymptotes, *a* its vertex, and *cag* its axis; if *adck* is a square and *he* is parallel to *ad*, and $cd = 1$, and $de = x$, then $be = \frac{1}{1+x}$. If also, $ef = 1$, $eg = 1 + x$, $eh = 1 + 2x + x^2$ etc. (the progression continued is $1 + 3x + 3xx + x^3$, $1 + 4x + 6x^2 + 4x^3 + x^4$, $1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5$ etc).

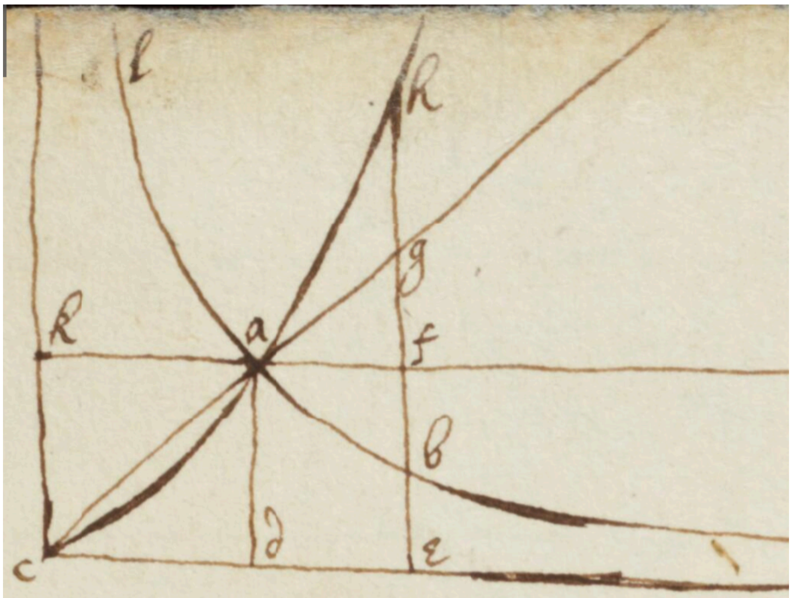


Figure 2: Newton's sketch of the first four curves in Equation 1.

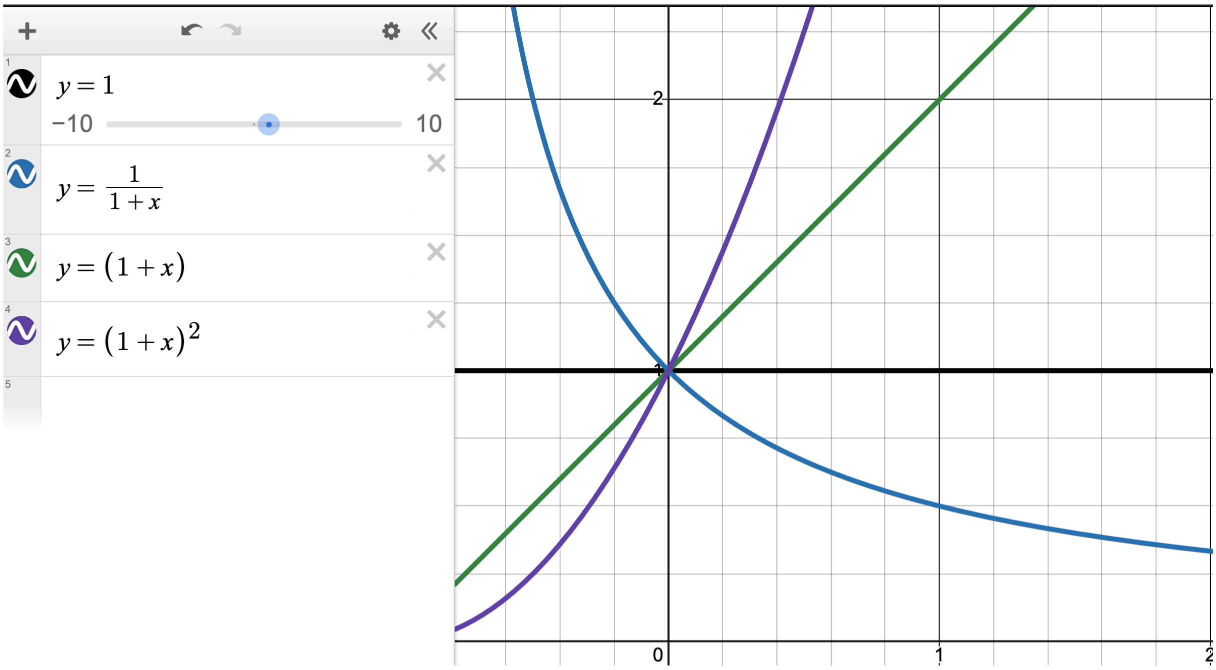


Figure 3: We have plotted[9] the first four curves in Equation 1. Compare with Figure 2.

Newton evaluates the integrals of all of the curves, marking with a $\star$ the one that cannot be found by applying the simple rule $\int x^n dx = \frac{x^{n+1}}{n+1} + C$.

> Then, shall the areas of those lines proceed in this progression. $* = adeb$, $x = adef$, $x + \frac{xx}{2} = adeg$, $adeh = x + \frac{2xx}{2} + \frac{x^3}{3}$, $x + \frac{3xx}{2} + \frac{3x^3}{3} + \frac{x^4}{4}$, $x + \frac{4xx}{2} + \frac{6x^3}{3} + \frac{4x^4}{4} + \frac{x^5}{5}$ etc. As in this table. In which the first area is also inserted.

For the table in question see Figure 4

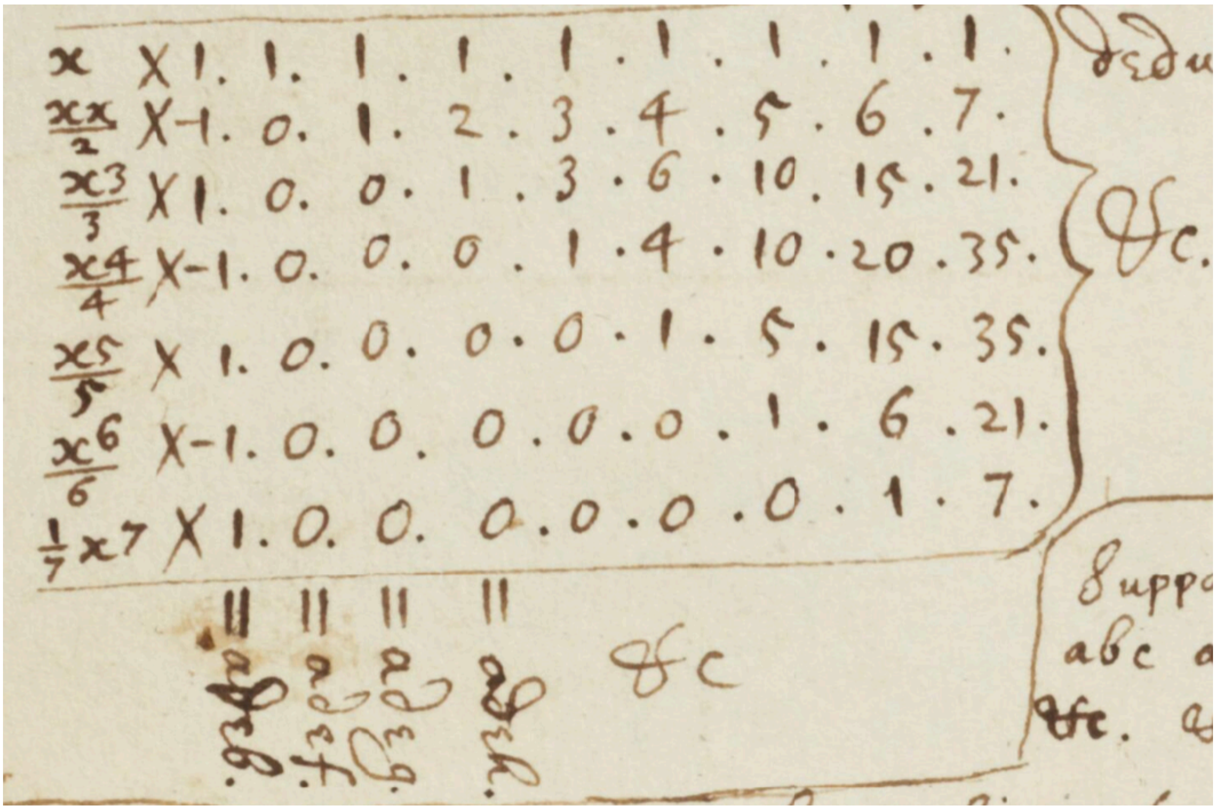
When we integrate the integrals in Equation 2 and express our expansion in terms of coefficients of terms of the form $\frac{x^n}{n}$ we see that the coefficients match those of 'Pascal's Triangle'[10]

[9] https://www.desmos.com/calculator/rnlmbugwau

[10] This eponym was not yet established in the decade following Blaise Pascal's death, so unsurprisingly Newton does not refer to Pascal.

$$
\begin{aligned}
\int \frac{1}{1+x}\,\mathrm{d}x &= \star \\
\int \mathrm{d}x &= \mathbf{1}x \\
\int (1+x)\,\mathrm{d}x &= \mathbf{1}x + \mathbf{1}\frac{x^2}{2} \\
\int (1+x)^2\,\mathrm{d}x &= \mathbf{1}x + \mathbf{2}\frac{x^2}{2} + \mathbf{1}\frac{x^3}{3} \\
\int (1+x)^3\,\mathrm{d}x &= \mathbf{1}x + \mathbf{3}\frac{x^2}{2} + \mathbf{3}\frac{x^3}{3} + \mathbf{1}\frac{x^4}{4} \\
\int (1+x)^4\,\mathrm{d}x &= \mathbf{1}x + \mathbf{4}\frac{x^2}{2} + \mathbf{6}\frac{x^3}{3} + \mathbf{4}\frac{x^4}{4} + \mathbf{1}\frac{x^5}{5} \\
\int (1+x)^5\,\mathrm{d}x &= \mathbf{1}x + \mathbf{5}\frac{x^2}{2} + \mathbf{10}\frac{x^3}{3} + \mathbf{10}\frac{x^4}{4} + \mathbf{5}\frac{x^5}{5} + \mathbf{1}\frac{x^6}{6}
\end{aligned}
\tag{2}
$$

In order to spot the pattern, Newton takes the coefficients in Equation 2, and arranges them in the following table (see Figure 4).



x	×	1.	1.	1.	1.	1.	1.	1.	1.	1.
xx/2	×	−1.	0.	1.	2.	3.	4.	5.	6.	7.
x3/3	×	1.	0.	0.	1.	3.	6.	10.	15.	21.
x4/4	×	−1.	0.	0.	0.	1.	4.	10.	20.	35.
x5/5	×	1.	0.	0.	0.	0.	1.	5.	15.	35.
x6/6	×	−1.	0.	0.	0.	0.	0.	1.	6.	21.
1/7 x7	×	1.	0.	0.	0.	0.	0.	0.	1.	7.

&c.

Figure 4: Each column represents the power series expansion of the integral of a polynomial. Successive rows represent the coefficients in front of terms of the form $\frac{x^n}{n}$.

Newton notes the Pascal's triangle pattern by saying "The sum of any figure and the figure above it is equal to the figure following it", which is clear by inspecting Table 1 (for example $1+1=2$, $3+1=4$, $0+0=0$ etc)

> The composition of which table may be deduced from hence; viz: The sum of any figure and the figure above it is equal to the figure following it.

We have transcribed Newton's table below:

		$\star$									
integral of $\rightarrow$ coefficient of $\downarrow$		$(1+x)^{-1}$	1	$\frac{1+x}{1}$	$\frac{(1+x)^2}{2}$	$\frac{(1+x)^3}{3}$	$\frac{(1+x)^4}{4}$	$\frac{(1+x)^5}{5}$	$\frac{(1+x)^6}{6}$	$\frac{(1+x)^7}{7}$
x	$\times$		1	1	1	1	1	1	1	1
$\frac{x^2}{2}$	$\times$		0	1	2	3	4	5	6	7
$\frac{x^3}{3}$	$\times$		0	0	1	3	6	10	15	21
$\frac{x^4}{4}$	$\times$		0	0	0	1	4	10	20	35
$\frac{x^5}{5}$	$\times$		0	0	0	0	1	5	15	35
$\frac{x^6}{6}$	$\times$		0	0	0	0	0	1	6	21
$\frac{x^7}{7}$	$\times$		0	0	0	0	0	0	1	7

Table 1: Transcription of Figure 4. The integral of $(1+x)^{-1}$ is the one to be found. We have inserted the missing sequence in Table 2 below.

1
-1
1
-1
1
-1
1

Table 2: Completed values for the missing column marked $\star$.

Newton concludes by stating the answer for $\star$:

> By which table it may appear that the area of the Hyperbola *adeb* is $x - \frac{xx}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \frac{x^6}{6} + \frac{x^7}{7} - \frac{x^8}{8} + \frac{x^9}{9} - \frac{x^{10}}{10}$ etc.

His 'extrapolated' value for $\star$, is:

$$\star = \mathbf{1}x - \mathbf{1}\frac{x^2}{2} + \mathbf{1}\frac{x^3}{3} - \mathbf{1}\frac{x^4}{4} + \mathbf{1}\frac{x^5}{5} + ... \tag{3}$$

How does Newton obtain this value? He does it by extrapolating the pattern in Table 1: the coefficient of x in the expansion of $\star$ must be 1, because all coefficients of x in all the other expansions are 1. Therefore the first empty cell in Table 1 must be a 1. The coefficient of $\frac{x^2}{2}$ must be -1, because this is required for the sum of any figure and the figure above it to be the figure next to it - i.e $1 - 1 = 0$. We continue this reasoning to get all the coefficients of the expansion of $\star$.

We may be tempted to call this a *discovery* of the integral of $\int \frac{1}{1+x} dx$ or a motivation of why this is a plausible value, rather than a *derivation* of it, because it crucially relies on the implicit assumption that this pattern is foolproof even for negative values of n. At this point, it would be a stretch to say that Newton has *applied* the binomial theorem to $\frac{1}{1+x}$. Rather, it seems like he has integrated $\frac{1}{1+x}$ by using the same style of reasoning that would later lead him to explicitly derive the binomial theorem (see Section 5).

Newton does not check his answer in this page of the manuscript, but the later Epistola Prior (Section 4) will show that he was aware he could obtain the same result through long division of $\frac{1}{1+x}$.

So far Newton has established that integrals of functions like $(1 + x)^n$ have regular patterns which can be *extrapolated* to $n = -1$, but he has not yet found a general rule for generating those patterns. He will be forced to do so in the next section, where his search for the integral of $(1 - x^2)^{\frac{1}{2}}$ will lead him to interpolate the integrals of the form $(1 - x^2)^n$ for $n = 0, 1, 2, 3$ to the cases $n = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$. As we shall see, Newton does this by first finding a general formula for the coefficients of $(1 - x^2)^n$ in terms of n, allowing him to plug in $n = \frac{1}{2}$. We shall see that these are exactly the binomial coefficients. As Steven Strogatz[11] recounts, Newton was likely inspired to use the method of interpolation by reading John Wallis' *Arithmetica Infinitorum* (1656)[12].

It may be asked whether Newton bothered to interpolate further than $n = -1$. Did he bother to look for the expansions of $(1 + x)^{-2}$ or $(1 + x)^{-3}$? Indeed he appears to have done so, for in MS Add. 3958:70v[13] we see a similar table to Figure 4 but extrapolated to $n = -4$.

2.2 Integral of Circle ($\int \sqrt{1 - x^2} dx$)

In 1655 John Wallis (1616-1703) had published his solution to the quadrature of a circle in his *Arithmetica Infinitorum* (1655), finding[14] the expression $\frac{\pi}{4} = \frac{2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdot 8 \cdot 8 * 10 \text{ etc}}{3 \cdot 3 \cdot 5 \cdot 5 \cdot 7 \cdot 7 \cdot 9 \cdot 9 \text{ etc}}$. Note that the denominator is composed of all the odd numbers twice, whereas the numerator is composed of all the even numbers twice except the first and last. Today this is known as the Wallis Product[15].

If we adopt a slightly more general notation, Wallis found $W_k \to \frac{\pi}{4}$ as $k \to \infty$, where:

[11] Steven Strogatz, *Infinite Powers: How Calculus Reveals the Secrets of the Universe*, First Mariner books edition (Mariner Books ; Houghton Mifflin Harcourt, 2020).

[12] John Wallis and Jacqueline A. Stedall, *The Arithmetic of Infinitesimals*, Sources and Studies in the History of Mathematics and Physical Sciences (Springer, 2004), §191, pp. 164-178.

[13] https://cudl.lib.cam.ac.uk/view/MS-ADD-03958/136

[14] Wallis and Stedall, *The Arithmetic of Infinitesimals*, p. 166.

[15] https://en.wikipedia.org/wiki/Wallis_product

$$W_k = \frac{2 \cdot 4}{3 \cdot 3} \times \frac{4 \cdot 6}{5 \cdot 5} \times \frac{6 \cdot 8}{7 \cdot 7} \times ... \times \frac{(k-1)(k+1)}{k} \tag{4}$$

Where k is odd.

We have tabulated W_k and the deviation with respect to $\frac{\pi}{4}$ in the table below:

	W_k	$W_k - \frac{\pi}{4}$
W_3	0.8888888889	0.10349072549144056
W_5	0.8533333333333334	0.06793516993588511
W_9	0.8255983875031494	0.04020022410570112
W_{33}	0.8187752603337018	0.03337709693625357
W_{99}	0.7893349223043913	0.003936758906943005

In Figure 5 we have plotted $W_k - \frac{\pi}{4}$ for increasing values of k:

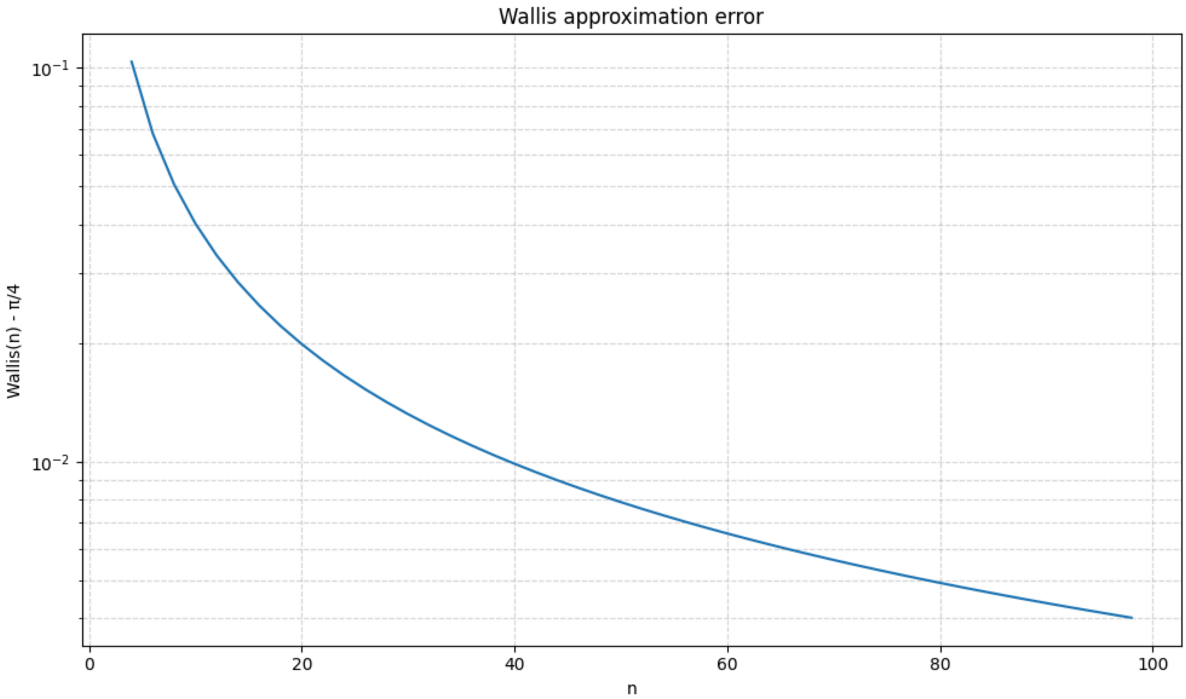


Figure 5: The difference between Wallis' product W_k and $\frac{\pi}{4}$ decreases with k

A twenty-two year old Newton read John Wallis' book in the winter of 1664-1665[16], making extensive notes. According to Strogatz[17], Wallis found his expression for W_k by the interpolation of the infinite series solution of integrals like $\int_0^1 (1-x^2)^n dx$ for integer n, 'filling the gaps' to find the values for half-odd integers n. But whereas Wallis' limits of integration were 0 and 1, Newton generalised his result by applying his method to arbitrary limits of integration, evaluating expressions like $\int_0^x (1-t^2)^n dt$.

[16]Wallis and Stedall, *The Arithmetic of Infinitesimals*, p. xxxi.

[17]Steven Strogatz, 'Strogatz_20newton_20series', 2019, https://www.stevenstrogatz.com/essays/2019/11/19/appendix-2-unpublished-from-infinite-powers.

As with the Hyperbola we covered in Section 2.1, Newton begins by drawing the curves of the form $(1-x^2)^m$ for $m = 0, \frac{1}{2}, 1, \frac{3}{2}, 2...$:

$$
\begin{aligned}
y &= 1 \\
y &= (1-x^2)^{\frac{1}{2}} \\
y &= (1-x^2) = 1-x^2 \\
y &= (1-x^2)^{\frac{3}{2}} \\
y &= (1-x^2)^2 = 1-2x^2+x^4 \\
y &= (1-x^2)^{\frac{5}{2}} \\
y &= (1-x^2)^3 = 1-3x^2+3x^4-x^6
\end{aligned}
\tag{5}
$$

> “ Suppose that *adck* is a Square, *abc* a circle, *age* a Parabola, etc. and that $de = x$ and *ad* is parallel to $fe = 1 = bc$. And that the progression in which the lines *fe*, *be*, *ge*, *he*, *ie*, *ne*, etc. proceeds is 1, $\sqrt{1-xx}$, $1-xx$, $\overline{1-xx}\sqrt{1-xx}$, $1-2xx+x^4$, $\overline{1-2xx+x^4}\sqrt{1-xx}$, $1-3xx+3x^4-x^6$, etc.

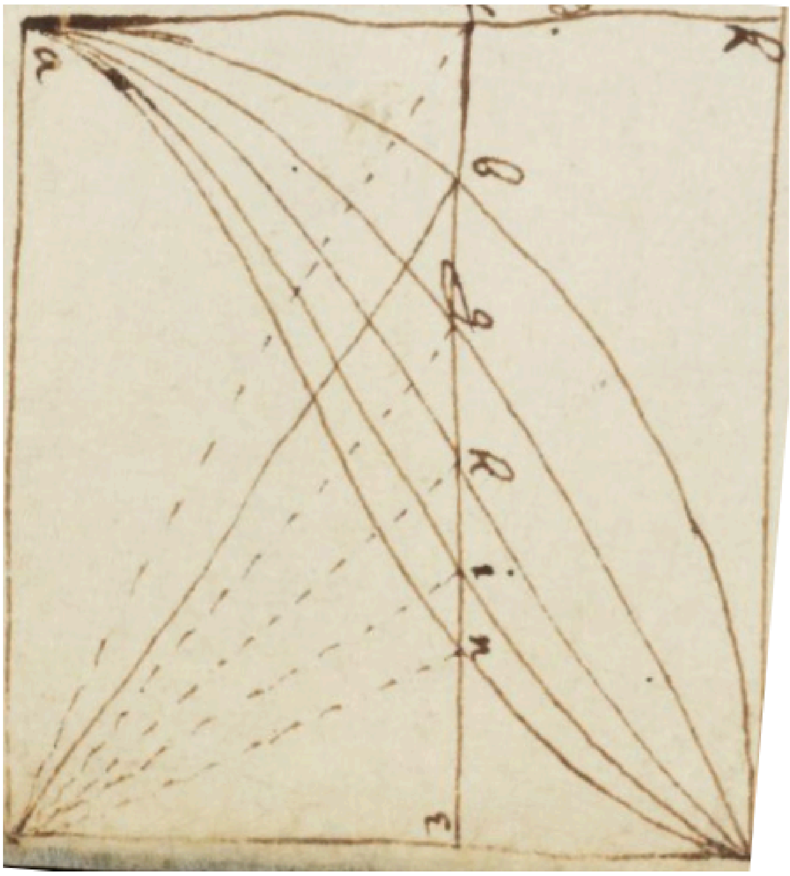


Figure 6: Newton's sketch of the curves in Equation 5. Since Newton sketches these with a vertical x-axis (see Figure 7, Left), we have rotated the screenshot of the manuscript. We have redrawn this graph in the right half of Figure 7. Newton provides an alternative version of the same drawing in Newton MS-ADD-04000 p. 39[18]

Looking at Figure 6, we see some dotted lines which don't seem to be used for anything in the ensuing calculation, but their lengths can be worked out by the Pythagorean theorem, and Newton was clearly thinking about them because he reserves the non-dashed line for connecting the origin to the curve he seeks to integrate - namely $(1-x^2)^{\frac{1}{2}}$. The curve $\sqrt{1-x^2}$ has the special property that a line connecting the origin to the point (x, y) has a length of 1 by the Pythagorean theorem.

[18]https://cudl.lib.cam.ac.uk/view/MS-ADD-04000/39

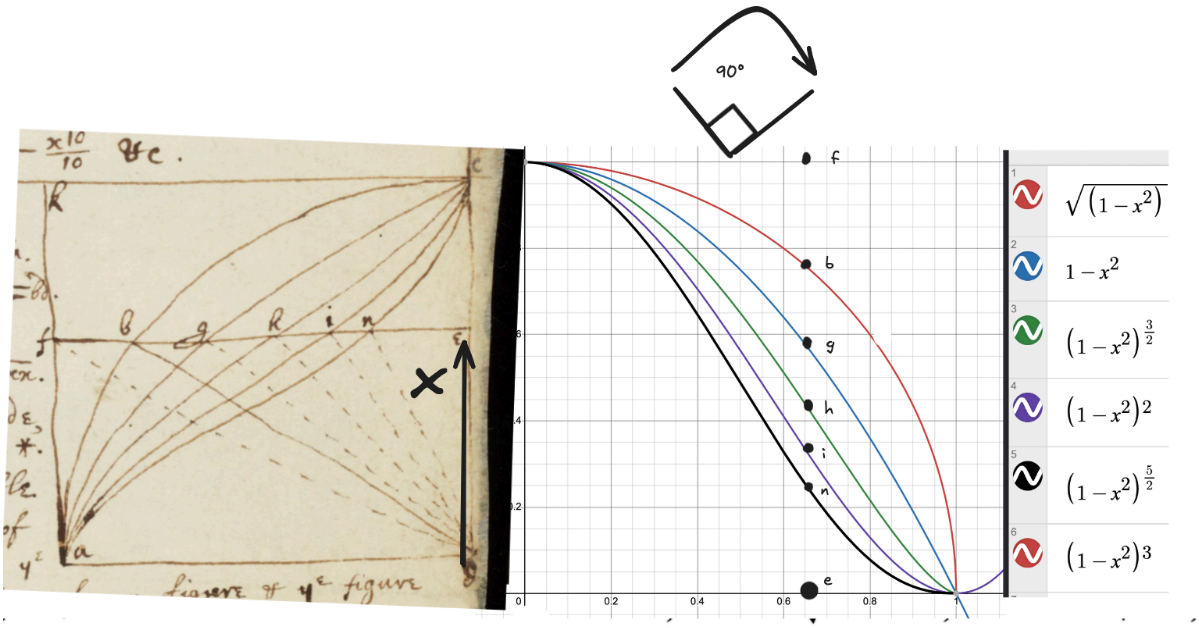


Figure 7: Left: Newton's original sketch, with a vertically oriented x-axis. Right: Modern sketch of the functions in Equation 5

Newton then proceeds to integrate the easy ones:

> Then will their areas *fade*, *bade*, *gade*, *hade*, *iade*, etc. be in this progression $x, *, x-\frac{xxx}{3}, *, x-\frac{2}{3}x^3+\frac{1}{5}x^5, *, x-\frac{3x^3}{3}+\frac{3x^5}{5}-\frac{x^7}{7}, *, x-\frac{4x^3}{3}+\frac{6x^5}{5}-\frac{4x^7}{7}+\frac{x^9}{9}$, etc: as in this table following in which the indeterminate terms are inserted.

In other words, he has just done these integrals:

$$\int (1-x^2)^0 \, \mathrm{d}x = \mathbf{1}x$$

$$\int (1-x^2)^{\frac{1}{2}} \, \mathrm{d}x = \star$$

$$\int (1-x^2) \, \mathrm{d}x = \mathbf{1}x + \mathbf{1}\left(-\frac{x^3}{3}\right)$$

$$\int (1-x^2)^{\frac{3}{2}} \, \mathrm{d}x = \star$$

$$\int (1-x^2)^2 \, \mathrm{d}x = \int (1-2x^2+x^4) \, \mathrm{d}x = \mathbf{1}x + \mathbf{2}\left(-\frac{x^3}{3}\right) + \mathbf{1}\frac{x^5}{5} \tag{6}$$

$$\int (1-x^2)^{\frac{5}{2}} \, \mathrm{d}x = \star$$

$$\int (1-x^2)^3 \, \mathrm{d}x = \int (1-3x^2+3x^4-x^6) \, \mathrm{d}x = \mathbf{1}x + \mathbf{3}\left(-\frac{x^3}{3}\right) + \mathbf{3}\frac{x^5}{5} + \mathbf{1}\left(-\frac{x^7}{7}\right)$$

$$\int (1-x^2)^{\frac{7}{2}} \, \mathrm{d}x = \star$$

$$\int (1-x^2)^4 \, \mathrm{d}x = \int (1-4x^2+6x^4-4x^6+x^8) \, \mathrm{d}x = \mathbf{1}x + \mathbf{4}\left(-\frac{x^3}{3}\right) + \mathbf{6}\frac{x^5}{5} + \mathbf{4}\left(-\frac{x^7}{7}\right) + \mathbf{1}\frac{x^9}{9}$$

Leaving as $\star$ those integrals which cannot be evaluated using the rule $\int x^n dx = \frac{x^{n+1}}{n+1}$. Again, we have arranged things such that Pascal's triangle coefficients are now in **bold**. Note the alternating signs between terms, and the fact that only odd powers of x are present.

> Then will their areas *fade*, *bade*, *gade*, *hade*, *iade*, etc. be in this progression $x, *, x - \frac{xxx}{3}, *, x - \frac{2}{3}x^3 + \frac{1}{5}x^5, *, x - \frac{3x^3}{3} + \frac{3x^5}{5} - \frac{x^7}{7}$, $*, x - \frac{4x^3}{3} + \frac{6x^5}{5} - \frac{4x^7}{7} + \frac{x^9}{9}$, etc: as in this table following in which the indeterminate terms are inserted.

n		-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{7}{2}$	4	$\frac{9}{2}$	5	$\frac{11}{2}$	6	type of progression
x	$\times$			1		1		1		1		1		1		1	constant
$-\frac{x^3}{3}$	$\times$			0		1		2		3		4		5		6	arithmetic
$\frac{x^5}{5}$	$\times$			0		0		1		3		6		10		15	triangular numbers
$-\frac{x^7}{7}$	$\times$			0		0		0		1		4		10		20	pyramidal numbers
$\frac{x^9}{9}$	$\times$			0		0		0		0		1		5		15	
$-\frac{x^{11}}{11}$	$\times$			0		0		0		0		0		1		6	
$\frac{x^{13}}{13}$	$\times$			0		0		0		0		0		0		1	

Table 3: Expansion coefficients of integrals of functions like $(1-x^2)^n$. Notice that the coefficients are exactly the same as the integrals of the form $(1+x)^n$.

Finding the series expansion of $\int \sqrt{1-x^2}dx$ boils down to finding the third empty column (corresponding to $n = \frac{1}{2}$) in Table 3. How do we go about this?

Again, Newton notices the same 'Pascal's triangle' pattern as in Section 2.1:

> The property of which table is that the sum of any figure and the figure above it is equal to the figure next after it save one.

Continuing the "Pascal's triangle" pattern allows us to fix the values for $n = -1$ much like we did in Section 2.1:

n		-1	$-\frac{1}{2}$
x	$\times$	1	
$-\frac{x^3}{3}$	$\times$	-1	
$\frac{x^5}{5}$	$\times$	1	
$-\frac{x^7}{7}$	$\times$	-1	
$\frac{x^9}{9}$	$\times$	1	
$-\frac{x^{11}}{11}$	$\times$	-1	
$\frac{x^{13}}{13}$	$\times$	1	

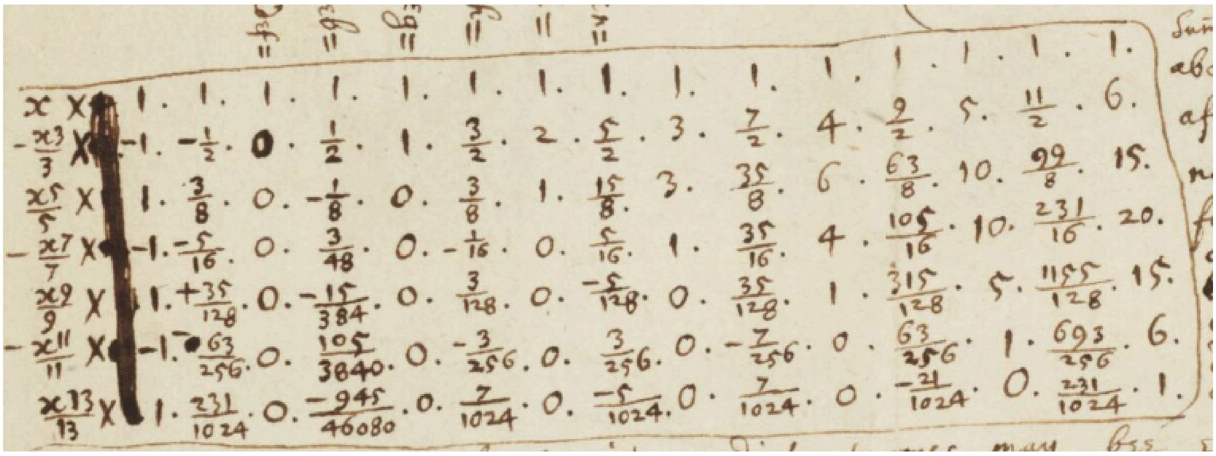
At this point it is clear that finding the value for $n = \frac{1}{2}$ can be done by interpolating the Table 3 for all half-odd integer values. How do we do this? For the first row (constants) and second row (arithmetic numbers) this seems rather simple: the values must be 1 and $\frac{1}{2}$ to match the constant and arithmetic series, respectively. It is interesting that Newton's account in his Epistola Prior[19] differs somewhat from the evidence from this particular manuscript. As we shall see in Section 5, Newton wrote to Leibniz that the general pattern for the binomial coefficients jumped out to him right then: Focusing on the second row in Table 4 (for

[19] Newton, 'Epistola Posterior (Letter to Gottfried Wilhelm Leibniz via Henry Oldenburg, October 24, 1676)'.

example, the '3' in the expansion $n = 3$), the rest of the terms in the column can be found by letting $m = 3$ and evaluating $\frac{m(m-1)}{2} = 3$ and $\frac{m(m-1)(m-2)}{2\times3} = 1$ respectively, until we reach $m\left(\frac{m-1}{2}\right)\left(\frac{m-2}{3}\right)\left(\frac{m-3}{4}\right) = 0$, after which we can stop. However, in this particular manuscript there is no trace of this formula in terms of m. Instead, it appears that Newton first found the intermediate terms by means of a finite difference table, and only later found a general formula for generating them.

> Also the numeral progressions are of these forms.
>
> | | | | |
> |---|---|---|---|
> | a | a | a | a |
> | b | $a+b$ | $2a+b$ | $3a+b$ |
> | c | $b+c$ | $a+2b+c$ | $3a+3b+c$ |
> | d | $c+d$ | $b+2c+d$ | $a+3b+3c+d$ |
> | e | $d+e$ | $c+2d+e$ | $b+3c+3d+e$ |
>
> Where the calculation of the intermediate terms may be easily performed.

It appears that Newton uses this table to infer the general form of the binomial coefficients, which allows him to infer the intermediate terms.



x	X	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
$-\frac{x^3}{3}$	X	-1.	$-\frac{1}{2}$.	0.	$\frac{1}{2}$.	1.	$\frac{3}{2}$.	2.	$\frac{5}{2}$.	3.	$\frac{7}{2}$.	4.	$\frac{9}{2}$.	5.	$\frac{11}{2}$.	6.
$\frac{x^5}{5}$	X	1.	$\frac{3}{8}$.	0.	$-\frac{1}{8}$.	0.	$\frac{3}{8}$.	1.	$\frac{15}{8}$.	3.	$\frac{35}{8}$.	6.	$\frac{63}{8}$.	10.	$\frac{99}{8}$.	15.
$-\frac{x^7}{7}$	X	-1.	$-\frac{5}{16}$.	0.	$\frac{3}{48}$.	0.	$-\frac{1}{16}$.	0.	$\frac{5}{16}$.	1.	$\frac{35}{16}$.	4.	$\frac{105}{16}$.	10.	$\frac{231}{16}$.	20.
$\frac{x^9}{9}$	X	1.	$+\frac{35}{128}$.	0.	$-\frac{15}{384}$.	0.	$\frac{3}{128}$.	0.	$-\frac{5}{128}$.	0.	$\frac{35}{128}$.	1.	$\frac{315}{128}$.	5.	$\frac{1155}{128}$.	15.
$-\frac{x^{11}}{11}$	X	-1.	$\frac{63}{256}$.	0.	$\frac{105}{3840}$.	0.	$-\frac{3}{256}$.	0.	$\frac{3}{256}$.	0.	$-\frac{7}{256}$.	0.	$\frac{63}{256}$.	1.	$\frac{693}{256}$.	6.
$\frac{x^{13}}{13}$	X	1.	$\frac{231}{1024}$.	0.	$\frac{-945}{46080}$.	0.	$\frac{7}{1024}$.	0.	$-\frac{5}{1024}$.	0.	$\frac{7}{1024}$.	0.	$-\frac{21}{1024}$.	0.	$\frac{231}{1024}$.	1.

Figure 8: Each column represents the integral of a polynomial of the form $\left(1-x^2\right)^m$ for $m = -1, \frac{1}{2}, 0, \frac{1}{2}, 1, \frac{3}{2}$. It may seem a little strange to see unreduced fractions like $\frac{3}{48}$ for $\frac{1}{16}$ or $\frac{105}{3840}$ for $\frac{21}{768}$ but we believe these confirm that they were probably calculated with the binomial theorem, i.e evaluating $\frac{m(m-1)(m-2)}{3!}$ for $m = \frac{1}{2}$ gives $\frac{3}{48}$ without cancellations.

> The area *abed* depends upon the 4th column 1, $\frac{1}{2}$, $-\frac{1}{8}$, $\frac{3}{48}$, etc. (which progression may be continued at pleasure by the help of this rule $\frac{0\times 1\times -1\times 3\times -5\times 7\times -9\times 11\times -13\times 15}{0\times 2\times 4\times 6\times 8\times 10\times 12\times 14\times 16\times 18}$ etc.) Whereby it may appear that, whatever the sine $de = x$ is, the area *abed* is $x - \frac{x^3}{6} - \frac{x^5}{40} - \frac{x^7}{112} - \frac{5x^9}{1152} - \frac{7x^{11}}{2816} - \frac{21x^{13}}{13312} - \frac{11x^{15}}{10240}$ etc. (and the area *afb* is $\frac{x^3}{6} + \frac{x^5}{40} + \frac{x^7}{112}$ etc.) Whereby also the area and angle *adb* may be found.

The rule $\frac{0\times 1\times -1\times 3\times -5\times 7\times -9\times 11\times -13\times 15}{0\times 2\times 4\times 6\times 8\times 14\times 16\times 18}$ is more accurately transcribed as:

$$\left(\frac{0}{0}\right)\left(\frac{1}{2}\right)\left(-\frac{1}{4}\right)\left(\frac{3}{6}\right)\left(-\frac{5}{8}\right)\left(\frac{7}{10}\right)\left(-\frac{9}{12}\right)\left(\frac{11}{14}\right)\left(-\frac{13}{16}\right)\left(\frac{15}{18}\right) \tag{7}$$

Where $\left(\frac{0}{0}\right)$ surprisingly evaluates to 1, $\left(\frac{0}{0}\right)\left(\frac{1}{2}\right)$ evaluates to $\frac{1}{2}$, $\left(\frac{0}{0}\right)\left(\frac{1}{2}\right)\left(-\frac{1}{4}\right)$ evaluates to $-\frac{1}{8}$ etc...

We can view this a special case of the binomial theorem, i.e calculating the coefficients $\binom{\frac{1}{2}}{0}$, $\binom{\frac{1}{2}}{1}$, $\binom{\frac{1}{2}}{2}$... where:

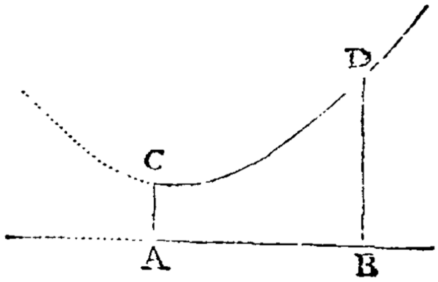
$$\begin{gathered}
\binom{\frac{1}{2}}{0} = 1 \\
\binom{\frac{1}{2}}{1} = \frac{1}{2} \\
\binom{\frac{1}{2}}{2} = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}-1\right)}{2} = -\frac{1}{8} \\
\binom{\frac{1}{2}}{3} = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}-1\right)\left(\frac{1}{2}-2\right)}{3 * 2} = \frac{1}{16} \\
\text{etc...}
\end{gathered} \tag{8}$$

The area *abed* is the area he set out to calculate, namely the area under the red curve on the right of Figure 7. The area afd is just $x - abed$. Since x is the sine of the angle adb, we can find the angle adb by taking the arcsine of x.

3 Root Extraction

In the following proof-of-concept extract[20], Newton shows that in principle we can obtain infinite series - by adapting the root extraction algorithm for decimal numbers to algebraic expressions.

[20]Newton, *Analysis by Means of Equations in an Infinite Number of Terms*, 1711

> ### *Exempla Radicem Extrahendo.*
>
> Si sit $\sqrt{aa+xx}=y$, Radicem sic extraho,
>
> $$aa+xx\ \Big(a+\frac{x^2}{2a}-\frac{x^4}{8a^3}+\frac{x^6}{16a^5}-\frac{5x^8}{128a^7}\ \&c.$$
>
> $$\frac{aa}{0}+xx$$
> $$\frac{xx+\frac{x^4}{4a^2}}{0-\frac{x^4}{4a^2}}$$
> $$-\frac{x^4}{4a^2}-\frac{x^6}{8a^4}+\frac{x^8}{64a^6}$$
> $$0+\frac{x^6}{8a^4}-\frac{x^8}{64a^6}$$
> $$+\frac{x^6}{8a^4}+\frac{x^8}{16a^6}-\frac{x^{10}}{64a^8}+\frac{x^{12}}{256a^{10}}$$
> $$0-\frac{5x^8}{64a^6}+\frac{x^{10}}{64a^8}-\frac{x^{12}}{256a^{10}}$$
> $$\&c.$$
>
> 
>
>
> Unde,

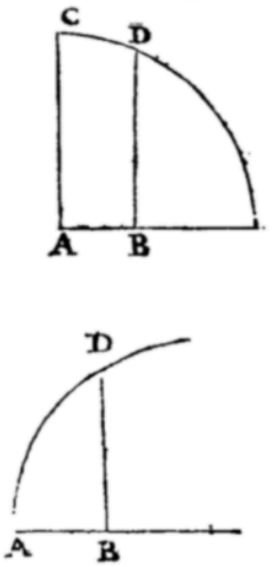
The first thing to note is that the left bracket has nothing to do with bracketing. Instead, it is supposed to separate the square of our answer from our answer:

$$\overbrace{aa+xx}^{\text{square of answer}}\ \Big(\underbrace{a+\frac{x^2}{2a}-\frac{x^4}{8a^3}+\frac{x^6}{16a^5}-\frac{5x^3}{128a^7}\quad\text{etc.}}_{\text{answer}}$$

The algorithm goes as follows: We start with guessing a number whose square will equal a^2. This is obviously a. Then we take our guess (a), square it (a^2), and take it away from our target answer a^2+x^2. The remainder is x^2. We take the remainder, divide it by $2a$, and add this value to our old guess (a), to produce our new guess ($a+\frac{x^2}{2a}$). Then repeat the process by squaring our new guess ($a^2+x^2+\frac{x^4}{4a^2}$) and taking this away from our target (still a^2+x^2). The remainder ($-\frac{x^4}{4a^2}$) is, like before, divided by 2a and added to our old guess ($a+\frac{x^2}{2a}$), creating a new guess ($a+\frac{x^2}{2a}-\frac{x^4}{8a^3}$). This process can continue forever, as the remainder will keep becoming smaller and smaller.

This algorithm is harder than standard long division because long division uses a constant divisor throughout the entire process, making each step a simple, linear repetition. In the root extraction algorithm, the divisor changes dynamically with every single iteration because you have to square the new, increasingly complex guess. This squaring action introduces higher-order polynomial terms (like x^4, x^6, and beyond), forcing you to manage exponentially larger numbers and do far more strenuous algebra just to calculate the next remainder. It is therefore *root extraction*, not *polynomial division*, that Newton identifies as a key procedure simplified by the binomial theorem (as we shall see in Section 4)

In the next step Newton takes this polynomial expansion of the curve, and integrates it term by term to get the area ABCD, thus squaring the hyperbola.

> PER ÆQUATIONES INFINITAS. 7
>
> Unde, pro Æquatione $\sqrt{aa+xx} = y$, nova producitur, viz. $y = a + \frac{x^2}{2a} - \frac{x^4}{8a^3} + \frac{x^6}{16a^5} - \frac{5x^8}{128a^7}$ &c. Et (per Reg. 2.) Area quæsita ABDC erit $= ax + \frac{x^3}{6a} - \frac{x^5}{40a^3} + \frac{x^7}{112a^5} - \frac{5x^9}{1152a^7}$ &c. Et hæc est Quadratura Hyperbolæ.
>
> 
>

So, the binomial theorem is not even necessary to square the hyperbola. We can generate power series algorithmically instead. But it is telling that algorithmic root extraction is not how Newton stumbled across the the power series for the hyperbola in the first place.

4 Epistola Prior (1676)

We next turn to the English translation of a letter, written in Latin, addressed to Henry Oldenburg but clearly intended to be read by Gottfried Wilhelm Leibniz (1646 - 1716). In it, Newton presents his version of the binomial theorem in recursive form, as well as the novel notation he has come up with (such as writing $\sqrt{a}$ as $a^{\frac{1}{2}}$, and $\frac{1}{a}$ as a^{-1}. We refer to the version in the textbook by Barrow-Green, Gray, and Wilson[21]

> Though the modesty of Mr Leibniz, in the extracts from his letter which you have lately sent me, pays great tribute to our countrymen for a certain theory of infinite series, about which there now begins to be some talk, yet I have no doubt that he has discovered not only a method for reducing any quantities whatever to such series, as he asserts, but also various shortened forms, perhaps like our own, if not even better. Since, however, he very much wants to know what has been discovered in this subject by the English, and since I myself fell upon this theory some years ago, I have sent you some of those things which occurred to me in order to satisfy his wishes, at any rate in part.
>
> Fractions are reduced to infinite series by division;

A simple example of reducing fractions to infinite series by division is Mercator's derivation of the series for $(1+x)^{-1}$ published in his *Logarithmotechnica*[22]:

[21] June Barrow-Green, Jeremy Gray, and Robin Wilson, *The History of Mathematics: A Source-Based Approach Vol 2* (2020), p. 86.

[22] Nicolaus Mercator, *Logarithmotechnia: Sive, Methodus Construendi Logarithmos Nova, Accurata, Et Facilis* (Guilielmi Godbid, 1668).

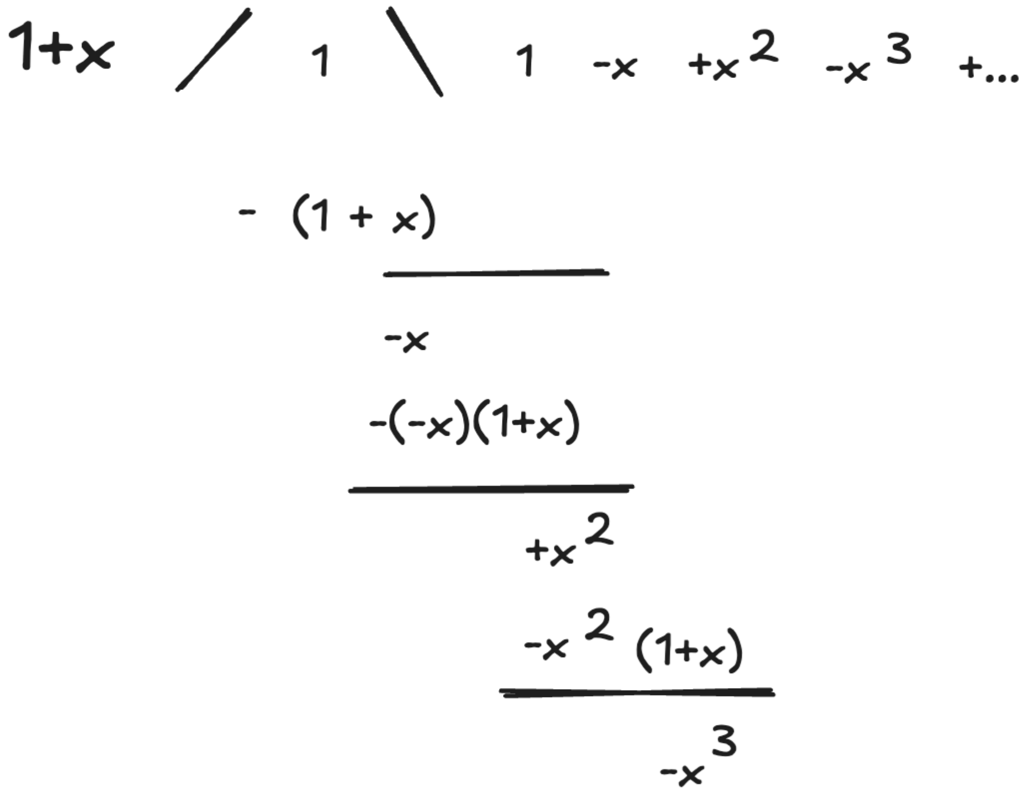


Figure 10: What Newton means when he says “fractions are reduced to infinite series by means of division”

> “ and quantities by extraction of the roots,

For an example of the root extraction algorithm for generating infinite series we refer the reader to our Section 3.

> “ by carrying out those operations in the symbols just as they are commonly carried out in decimal numbers.

Newton thus viewed power series generation as an algorithmic process much akin to the algorithms for long division or taking square roots. To quote Steven Strogatz:

> Newton viewed power series as a natural generalization of infinite decimals. An infinite decimal, after all, is nothing but an infinite series of powers of 10 and 1/10. The digits in the number tell us how much of each power of 10 or 1/10 to mix in. For example, the number $\pi = 3.14\ldots$ corresponds to this particular mix: $3.14... = 3 \times (10)^0 + 1 \times \left(\frac{1}{10}\right)^1 + 4 \times \left(\frac{1}{10}\right)^2 + ...$
>
> — Strogatz, [p. 189]

In light of this understanding, it is not surprising that Newton considered the manual algorithms for long division and root extraction to lie at the *foundation* of generating algebraic power series:

> These are the foundations of these reductions: but extractions of roots are much shortened by this theorem,
>
> $$(P + PQ)^{m/n} = P^{m/n} + \frac{m}{n}AQ + \frac{m-n}{2n}BQ + \frac{m-2n}{3n}CQ + \frac{m-3n}{4n}DQ + etc.$$
>
> where $P + PQ$ signifies the quantity whose root or even any power, or the root of a power, is to be found; P signifies the first term of that quantity, Q the remaining terms divided by the first, and m/n the numerical index of the power of $P + PQ$, whether that power is integral or (so to speak) fractional, whether positive or negative.

We must keep in mind that manually working out $(1-x^2)^{\frac{1}{2}}$ through the "extraction of roots" algorithm is a much more laborious process than working out $(1+x)^{-1}$ through long division, therefore Newton claims it is the former that is much shortened by the binomial theorem. Of course, there are no algorithmic approaches for expressions like $(1-x^2)^{\frac{7}{2}}$, rendering the binomial theorem absolutely necessary for more complicated exponents.

Now that we have seen Newton's statement of the binomial theorem, let us see how it 'maps on' to our familiar, modern one, namely:

$$(x+y)^r = \binom{r}{0}x^r + \binom{r}{1}y^1x^{r-1} + ... \qquad \text{where:} \quad \binom{r}{k} = \frac{r(r-1)...(r-k+1)}{k!} \tag{9}$$

We start by writing the first few terms of the modern one:

$$\begin{aligned}(x+y)^r &= \binom{r}{0}x^{r-0}y^0 + \binom{r}{1}x^{r-1}y^1 + \binom{r}{2}x^{r-2}y^2 + \binom{r}{3}x^{r-3}y^3 + ... \\ &= x^r + rx^{r-1}y + \frac{r(r-1)}{2}x^{r-2}y^2 + \frac{r(r-1)(r-2)}{3 \times 2}x^{r-3}y^3 + ...\end{aligned} \tag{10}$$

Anticipating that we want $Q = \frac{y}{x}$ and $P = x$ to appear, we create groupings of $\frac{y}{x}$:

$$\left(x+x\frac{y}{x}\right)^r = x^r + rx^r\left(\frac{y}{x}\right) + \frac{r(r-1)}{2}x^r\left(\frac{y}{x}\right)^2 + \frac{r(r-1)(r-2)}{3\times 2}x^r\left(\frac{y}{x}\right)^3 + ... \tag{11}$$

now we note that our r is actually Newton's fraction $\frac{m}{n}$:

$$\left(x+x\frac{y}{x}\right)^{\frac{m}{n}} = x^{\frac{m}{n}} + \frac{m}{n}x^{\frac{m}{n}}\left(\frac{y}{x}\right) + \frac{\frac{m}{n}\left(\frac{m}{n}-1\right)}{2}x^{\frac{m}{n}}\left(\frac{y}{x}\right)^2 + \frac{\left(\frac{m}{n}\right)\left(\frac{m}{n}-1\right)\left(\frac{m}{n}-2\right)}{3\times 2}x^{\frac{m}{n}}\left(\frac{y}{x}\right)^3 + \tag{12}$$

We can simplify some of the fractions by multiplying through by $\frac{n}{n}$:

$$\left(x+x\frac{y}{x}\right)^{\frac{m}{n}} = x^{\frac{m}{n}} + \frac{m}{n}x^{\frac{m}{n}}\left(\frac{y}{x}\right) + \frac{\frac{m}{n}(m-n)}{2n}x^{\frac{m}{n}}\left(\frac{y}{x}\right)^2 + \frac{\left(\frac{m}{n}\right)(m-n)\left(\frac{m}{n}-2\right)}{(3\times 2)n}x^{\frac{m}{n}}\left(\frac{y}{x}\right)^3 + \tag{13}$$

Now we are ready to 'connect' with Newton's formula. Clearly, $P = x$, $Q = \frac{y}{x}$. The first term of the expansion, A, is $x^{\frac{m}{n}}$ (or $P^{\frac{m}{n}}$). Let us write everything in terms of P, Q and A:

$$(P+PQ)^{\frac{m}{n}} = \underbrace{P^{\frac{m}{n}}}_{\text{A}} + \underbrace{\frac{m}{n}AQ}_{\text{B}} + \left(\frac{m}{n}AQ\right)\frac{m-n}{2n}Q + \frac{\frac{m}{n}AQ(m-n)\left(\frac{m}{n}-2\right)}{(3\times 2)n}Q^2 + ... \tag{14}$$

In the above formula, we have spotted that the term $x^{\frac{m}{n}}$ occurs in every term, so we have replaced it by A. Now, $B = \frac{m}{n}AQ$ is the second term (by definition), and we also spot that all the remaining terms have a B hiding in them.

$$(P+PQ)^{\frac{m}{n}} = \underbrace{P^{\frac{m}{n}}}_{\text{A}} + \underbrace{\frac{m}{n}AQ}_{\text{B}} + \underbrace{B\frac{m-n}{2n}Q}_{\text{C}} + \underbrace{\frac{BQ(m-n)\left(\frac{m}{n}-2\right)}{(3\times 2)n}Q}_{\text{D}} + ... \tag{15}$$

Similarly, there is a C hiding in the D term, etc...

$$(P+PQ)^{\frac{m}{n}} = \underbrace{P^{\frac{m}{n}}}_{\text{A}} + \underbrace{\frac{m}{n}AQ}_{\text{B}} + \underbrace{\frac{m-n}{2n}BQ}_{\text{C}} + \underbrace{\frac{m-2n}{3n}CQ}_{\text{D}} + ... \tag{16}$$

And thus we have shown that a few algebraic manipulations takes us from the modern statement of the binomial theorem to Newton's recursive one.

Newton's preference for a recursive formula may have been because it mimicked the process of calculating roots using the 'extraction of roots' algorithm (terms are always calculated in order). It shows that he never had any use for calculating terms *out of order*, which is an advantage of the modern statement of the binomial theorem.

We might also add that multiplication[23] is an 'expensive' operation, so it makes sense to keep track of the terms as we are going and only multiply when we need to.

In treating fractional and integer powers in the same equation, Newton takes the opportunity to not only introduce his theorem but also to introduce an arguably greater contribution to mathematics: writing $a^{\frac{1}{2}}$ for $\sqrt{a}$ and a^{-1} for $\frac{1}{a}$. John Wallis had considered fractional

[23]Indeed, the expensiveness of multiplication (compared with addition) is what made the arrival of logarithms so attractive in the seventeenth century

exponents[24], but he seemed to have always written them like $\sqrt[4]{x}^3$. Newton takes explicit credit for his innovation:

> For as analysts, instead of aa, aaa, etc., are accustomed to write a^2, a^3, etc., so instead of $\sqrt{a}, \sqrt{a^3}, \sqrt{c : a^5}$, etc.[8] I write $a^{1/2}, a^{3/2}, a^{5/3}$, and instead of $1/a, 1/aa, 1/a^3$, I write a^{-1}, a^{-2}, a^{-3}.

He says that *just like* analysts have recently extended the notation of mathematics to include things like a^2 for aa, he goes a *step further* and uses things like $a^{\frac{3}{4}}$ and a^{-1} to denote $\left(\sqrt[4]{a}\right)^3$ and $\frac{1}{a}$. This notation allows him to treat both fractional and integer powers with one formula.

Newton's claim to novelty is undermined a little bit by the translation. In Latin[25], it is more clear that he is signalling an original contribution in the second part of this sentence, as he explicitly uses the word "I" (*ego*) as opposed to relying on the conjugation of the word "write" (*scribo*).

He next gives an example of the notation:

> And so for $\frac{aa}{\sqrt{c : (a^3 + bbx)}}$ I write $aa(a^3 + bbx)^{-1/3}$ and for
>
> $$\frac{aab}{\sqrt{c : \{(a^3 + bbx)(a^3 + bbx)\}}}$$
>
> I write $aab(a^3 + bbx)^{-2/3}$:

The c: means cube root.

[24]David Dennis and Jere Confrey, 'The Creation of Continuous Exponents: A Study of the Methods and Epistemology of John Wallis', in *CBMS Issues in Mathematics Education*, vol. 6, *CBMS Issues in Mathematics Education*, ed. James Kaput et al. (American Mathematical Society, 1996), https://doi.org/10.1090/cbmath/006/02.

[25]see https://cudl.lib.cam.ac.uk/view/MS-ADD-03977/63

> in which last case, if $(a^3+bbx)^{-2/3}$ is supposed to be $(P+PQ)^{m/n}$ in the Rule, then P will be equal to a^3, Q to bbx/a^3, m to -2, and n to 3.

His first example of using the binomial theorem involves the fractional power $\frac{2}{3}$.

> Finally, for the terms found in the quotient in the course of the working I employ A, B, C, D, etc., namely, A for the first term, $P^{m/n}$; B for the second term, m/nAQ; and so on.

Newton points out that the terms A, B, C, D are simply consecutive terms of the expansion. A more verbose way of stating Newton's binomial theorem may be:

$$(P+PQ)^{\frac{m}{n}} = A+B+C+D+E+.... \tag{17}$$

where:

$$\begin{aligned} A &= P^{\frac{m}{n}} \\ B &= \frac{m}{n}AQ \\ C &= \frac{m-n}{2n}BQ \\ D &= \frac{m-2n}{3n}CQ \\ &\dots \end{aligned} \tag{18}$$

If we allow ourselves to use now-standard notation for recursive formulae, we may express the binomial theorem as follows:

$$\sum_{r=0}^{\infty} T_r \tag{19}$$

Where:

$$T_0 = P^{\frac{m}{n}} \tag{20}$$

and:

$$T_r = \frac{m-(r-1)n}{rn}T_{r-1} \tag{21}$$

4.1 Finding $\sqrt{10}$ using Newton's recursive binomial theorem

Newton *Epistola Prior* was concerned with roots of algebraic expressions, not numbers. Nevertheless, let's give a more concrete example and see how to find $\sqrt{10}$ using Newton's recursive binomial theorem:

$$m = 1 \tag{22}$$

$$n = 2 \tag{23}$$

$$P + PQ = 10 \tag{24}$$

If we let $P = 9$ for the sake of convenience then we can write $P + PQ = 10 = 9 + 1 = 9 + 9 \times \left(\frac{1}{9}\right)$, so:

$$\begin{aligned} Q &= \frac{1}{9} \\ P &= 9 \end{aligned} \tag{25}$$

Then

$$\begin{aligned} (P + PQ)^{\frac{m}{n}} &= \left(9 + 9 \times \left(\frac{1}{9}\right)\right)^{\frac{1}{2}} \\ &= 9^{\frac{1}{2}} + \left(\frac{1}{2}\right)AQ + \frac{1-2}{(2)(2)}(BQ) + \frac{1-2(2)}{(3)(2)}(CQ) + \frac{1-3(2)}{(4)(2)}(DQ) + ... \end{aligned} \tag{26}$$

and using the fact that $Q = \frac{1}{9}$:

$$\begin{gathered} (P + PQ)^{\frac{m}{n}} = \\ 3 + \left(\frac{1}{2}\right)A\left(\frac{1}{9}\right) + \frac{1-2}{(2)(2)}\left(B\left(\frac{1}{9}\right)\right) + \frac{1-2(2)}{(3)(2)}\left(C\left(\frac{1}{9}\right)\right) + \frac{1-3(2)}{(4)(2)}\left(D\left(\frac{1}{9}\right)\right) \end{gathered} \tag{27}$$

$$A = 3 \tag{28}$$

$$B = \frac{1}{2} \times 3 \times \frac{1}{9} = \frac{1}{6} \tag{29}$$

$$C = \frac{1-2}{4} \times \frac{1}{6} \times \frac{1}{9} = -\frac{1}{216} \tag{30}$$

$$D = \frac{1-4}{6} \times -\frac{1}{216} \times \frac{1}{9} = \frac{1}{3888} \tag{31}$$

$$E = \frac{1-6}{8} \times \frac{1}{3888} \times \frac{1}{9} = -\frac{5}{279936} \tag{32}$$

Indeed

$$(A + B + C + D + E)^2 = 9.99999188742702 \tag{33}$$

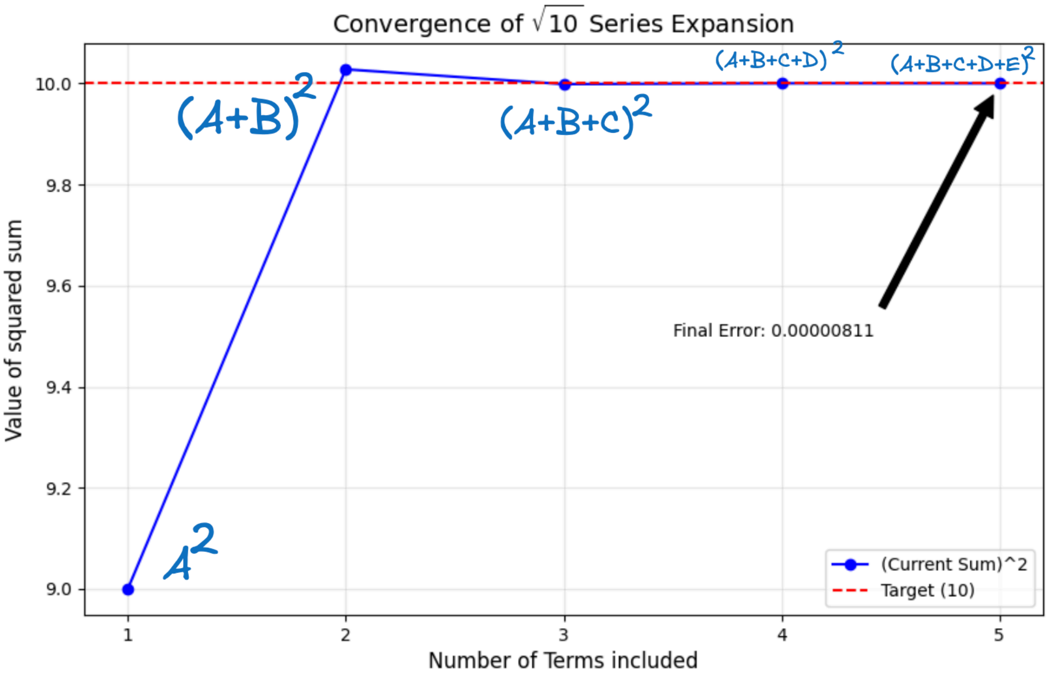


Figure 11: Graph to show rapid conversion of the binomial expansion of square root of 10

Newton next gives us some algebraic examples, which also show the application of his new notation for roots. The first example is a root extraction.

> *Example 1*
>
> $$\sqrt{(c^2+x^2)} \text{ or } (c^2+x^2)^{1/2} = c + \frac{x^2}{2c} - \frac{x^4}{8c^3} + \frac{x^6}{16c^5} - \frac{5x^8}{128c^7} + \frac{7x^{10}}{256[c]^9} + \textit{ etc.}$$
>
> For in this case $P = c^2, Q = x^2/c^2, m = 1, n = 2,\ A\,(= P^{m/n} = (cc)^{1/2}) = c$, $B\,(= (m/n)AQ) = x^2/2c$, $C\left(= \frac{m-n}{2n}BQ\right) = -\frac{x^4}{8c^3}$; and so on.

The transcription by Stedall has a typo as $Q = \frac{c^4x-x^5}{c^5}$, in the first case, not $\frac{(c^4x-x)^5}{c^5}$. We should let $P = c^2$ if $\frac{x}{c} < 1$, otherwise we should let $P = x^2$.

> *Example 2*
>
> $$\sqrt[5]{(c^5 + c^4x - x^5)},$$
>
> i.e.
>
> $$(c^5 + c^4x - x^5)^{1/5} = c + \frac{c^4x - x^5}{5c^4}[+]\frac{-2c^8x^2 + 4c^4x^6 - 2x^{10}}{25c^9} + etc.$$
>
> as will be evident on substituting 1 for m, 5 for n, c^5 for P and $(c^4x{-}x)^5/c^5$ for Q, in the rule quoted above. Also $-x^5$ can be substituted for P and $(c^4x + c^5)/(-x^5)$ for Q. The result will then be
>
> $$\sqrt[5]{(c^5 + c^4x - x^5)} = -x + \frac{c^4x + c^5}{5x^4} + \frac{2c^8x^2 + 4c^9x + [2]c^{10}}{25x^9} + etc.$$
>
> The first method is to be chosen if x is very small, the second if it is very large.

In the second example, we use a fifth root, for which there is no known algorithm for generating power series expansions (like there is for square roots - see Section 3), rendering the binomial theorem indispensable for generating the the power series in this case. Again, there is a typo in the Stedall translation as we are substituting $\frac{c^4x-x^5}{c^5}$ for Q, not $\frac{(c^4x-x)^5}{c^5}$.

5 Epistola Posterior (1676)

This is first part of the second letter[26] Newton sent to Leibniz (via Oldenburg) about the infinite series he had found. In this letter, Newton appears explains how he was inspired by the work of John Wallis to use interpolation to find infinite series solutions for the areas under curves - a process we first saw in his manuscript in Section 2.

> I can hardly tell with what pleasure I have read the letters of those very distinguished men Leibniz and Tschirnhaus.

Ehrenfried Walther von Tschirnhaus (1651 - 1708) was a fellow mathematician and likely inventor of Porcelain[27]

[26]Transcribed from latin in *The History of Mathematics: A Source-Based approach, Volume 2*, Barrow-Green et al.

[27]C.M Quieroz & S. Agothopoulos, 2006 https://arxiv.org/abs/physics/0601111

> I can hardly tell with what pleasure I have read the letters of those very distinguished men Leibniz and Tschirnhaus. Leibniz's method for obtaining convergent series is certainly very elegant, and it would have sufficiently revealed the genius of its author, even if he had written nothing else. But what he has scattered elsewhere throughout his letter is most worthy of his reputation — it leads us also to hope for very great things from him. The variety of ways by which the same goal is approached has given me the greater pleasure, because three methods of arriving at series of that kind had already become known to me, so that I could scarcely expect a new one to be communicated to us. One of mine I have described before; I now add another, namely, that by which I first chanced on these series — for I chanced on them before I knew the divisions and extractions of roots which I now use.

Newton intends to do two things. First, to show a second method of generating infinite series. And second, to show how he arrived at the binomial theorem. By 1676, Newton considered the generation of infinite series to be a fundamentally arithmetic algorithm akin to operations on infinite decimals, but as we see in this letter (and as we saw in Section 2) this is not how he first discovered these series. This illustrates that in mathematics the context of discovery can be different from the context of justification.

> And an explanation of this will serve to lay bare, what Leibniz desires from me, the basis of the theorem set forth near the beginning of the former letter.

> At the beginning of my mathematical studies, when I had met with the works of our celebrated Wallis, on considering the series by the intercalation of which he himself exhibits the area of the circle and the hyperbola, the fact that, in the series of curves whose common base or axis is x and the ordinates
>
> $$(1-x^2)^{0/2}, (1-x^2)^{1/2}, (1-x^2)^{2/2}, (1-x^2)^{3/2}, (1-x^2)^{4/2}, (1-x^2)^{5/2}, etc,$$
>
> if the areas of every other of them, namely
>
> $$x^3, x-\frac{1}{3}x^3, x-\frac{2}{3}x^3+\frac{1}{5}x^5, x-\frac{3}{3}x^3+\frac{3}{5}x^5-\frac{1}{7}x^7, etc.$$
>
> could be interpolated, we should have the areas of the intermediate ones, of which the first $(1-x^2)^{1/2}$ is the circle: in order to interpolate these series I noted that in all of them the first term was x and the second terms $\frac{0}{3}x^3$, $\frac{1}{3}x^3$, $\frac{2}{3}x^3$, $\frac{3}{3}x^3$, etc., were in arithmetical progression, and hence that the first two terms of the series to be intercalated ought to be $x-\frac{1}{3}(\frac{1}{2}x^3)$, $x-\frac{1}{3}(\frac{3}{2}x^3)$, $x-\frac{1}{3}(\frac{5}{2}x^3)$, etc.

It is likely that Newton obtained a copy of Wallis' *Arithmetica Infinitorum* from Isaac Barrow[28]. What Newton is saying here is that he got inspired by the work of Wallis to look at a set of functions, for the area was known for n even. He wanted to see if he could find a pattern to then also find the area for the functions for odd n. Again there appears to be a typo in the translation as the first term ought to be x, not x^3.

> To intercalate the rest I began to reflect that the denominators $1, 3, 5, 7$, etc., were in arithmetical progression, so that the numerical coefficients of the numerators only were still in need of investigation.

Here Newton is describing how he deduced the first two terms for the series expansion of the area of $(1-x^2)^{\frac{n}{2}}$ where $n \in \mathbb{N}$. When n is even, the expression has a finite expansion, and the areas are easily found by straightforward integration:

he could expand the area of the function and write down its terms, which allowed him to see a pattern emerge.

[28]June Barrow-Green, Jeremy Gray, and Robin Wilson, *The History of Mathematics: A Source-Based Approach Vol 2*, p. 53.

$$
\begin{gathered}
\int \left(1-x^2\right)^0 \mathrm{d}x = \mathbf{1}x \\
\int \left(1-x^2\right)^{\frac{1}{2}} \mathrm{d}x = \star \\
\int (1-x^2) \,\mathrm{d}x = \mathbf{1}x + \mathbf{1}\left(-\frac{x^3}{3}\right) \\
\int \left(1-x^2\right)^{\frac{3}{2}} \mathrm{d}x = \star \\
\int \left(1-x^2\right)^2 \mathrm{d}x = \int (1-2x^2+x^4) \,\mathrm{d}x = \mathbf{1}x + \mathbf{2}\left(-\frac{x^3}{3}\right) + \mathbf{1}\frac{x^5}{5} \\
\int \left(1-x^2\right)^{\frac{5}{2}} \mathrm{d}x = \star \\
\int \left(1-x^2\right)^3 \mathrm{d}x = \int (1-3x^2+3x^4-x^6) \,\mathrm{d}x = \mathbf{1}x + \mathbf{3}\left(-\frac{x^3}{3}\right) + \mathbf{3}\frac{x^5}{5} + \mathbf{1}\left(-\frac{x^7}{7}\right) \\
\int \left(1-x^2\right)^{\frac{7}{2}} \mathrm{d}x = \star \\
\int \left(1-x^2\right)^4 \mathrm{d}x = \int (1-4x^2+6x^4-4x^6+x^8) \,\mathrm{d}x = \mathbf{1}x + \mathbf{4}\left(-\frac{x^3}{3}\right) + \mathbf{6}\frac{x^5}{5} + \mathbf{4}\left(-\frac{x^7}{7}\right) + \mathbf{1}\frac{x^9}{9}
\end{gathered}
\tag{34}
$$

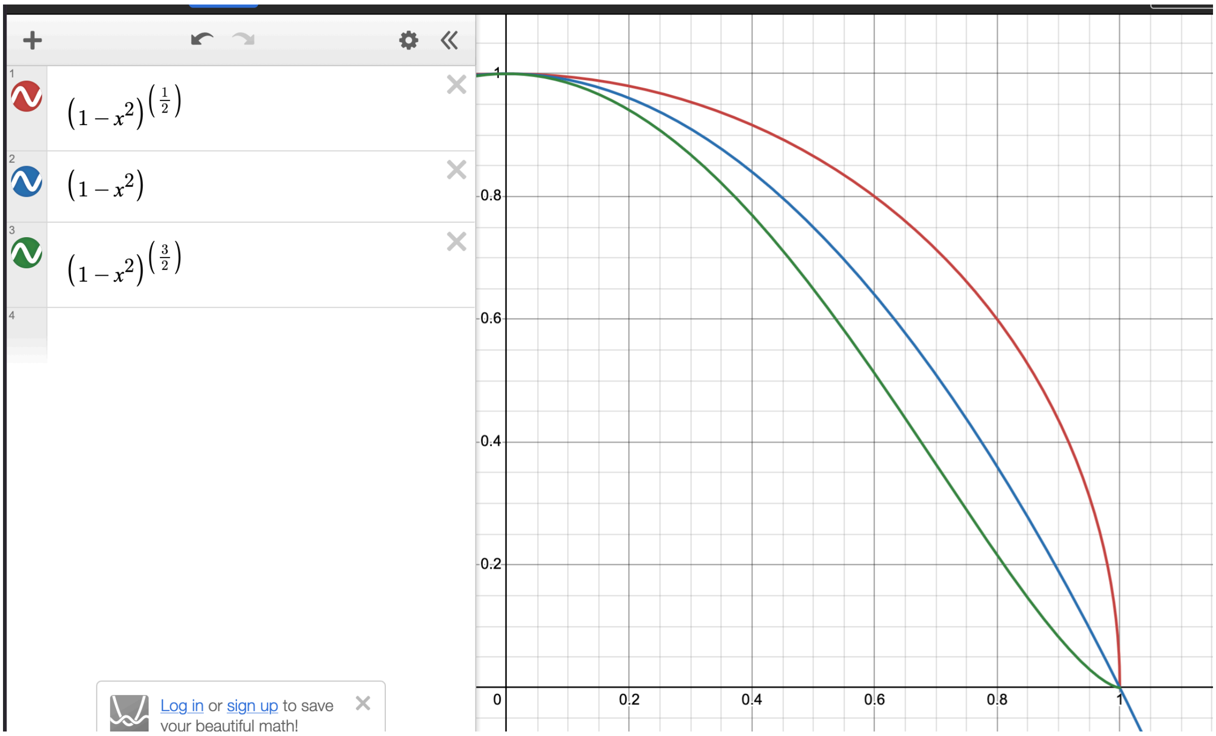


Figure 12: $(1-x^2)^{\frac{n}{2}}$ plotted for $n = \frac{1}{2}, 1, \frac{3}{2}$

> But in the alternately given areas these were the figures of powers of the number 11, namely of these, $11^0, 11^1, 11^2, 11^3, 11^4$, that is, first 1; then $1, 1$; thirdly, $1, 2, 1$; fourthly $1, 3, 3, 1$; fifthly $1, 4, 6, 4, 1$, etc.

This numerical pattern corresponds to what is now called Pascal's triangle (i.e., the binomial coefficients). However, for powers of the number 11, this pattern is valid up to 11^5 or higher powers due to multi-digit numbers, but not if you carry over the tens digit to the next place value:

$$\begin{gathered}\text{Row } 5: 1, 5, 10, 10, 5, 1 \\ \text{Carry over } 1, (5+1), (0+1), 0, 5, 1 = 161051 = 11^5\end{gathered} \tag{35}$$

This is just a different way of saying "the sum of the figure and the figure above it is equal to the figure next to it" - which we saw Newton do in his Manuscript in Section 2.

> $1, 2, 1$; fourthly $1, 3, 3, 1$; fifthly $1, 4, 6, 4, 1$, etc. And so I began to inquire how the remaining figures in these series could be derived from the first two given figures, and I found that on putting m for the second figure, the rest would be produced by continual multiplication of the terms of this series,
>
> $$\frac{m-0}{1} \times \frac{m-1}{2} \times \frac{m-2}{3} \times \frac{m-3}{4} \times \frac{m-4}{5} \times etc.$$

Newton here describes the recursive rule that generates the numbers for the numerators. He next gives an example which helps clear it up a bit more.

> For example, let $m = 4$, and $4 \times \frac{1}{2}(m-1)$, that is 6 will be the third term, and $6 \times \frac{1}{3}(m-2)$, that is 4 the fourth, and $4 \times \frac{1}{4}(m-3)$, that is 1 the fifth, and $1 \times \frac{1}{5}(m-4)$, that is 0 the sixth, at which term in this case the series stops.

Here Newton gives an example to find the rest of the binomial coefficients for $(a+b)^4$ once you know the first two, i.e: $(a+b)^4 = (a + 4a^3b + ...)$. The second coefficient is:

$$\frac{m-0}{1} = \frac{4-0}{1} = 4 \tag{36}$$

The third coefficient is:

$$\frac{m-0}{1} \times \frac{m-1}{2} = \frac{4-0}{1} \times \frac{4-1}{2} = 6 \tag{37}$$

The fourth coefficient is:

$$\frac{m-0}{1} \times \frac{m-1}{2} \times \frac{m-2}{3} = \frac{4-0}{1} \times \frac{4-1}{2} \times \frac{4-2}{3} = 4 \times \frac{3}{2} \times \frac{2}{3} = 4 \tag{38}$$

And the fifth (last) coefficient is 1.

So by changing the method by which he generates the binomial coefficients in known cases, he has come up with a formula that may be applied for non-integral exponents:

> Accordingly, I applied this rule for interposing series among series, and since, for the circle, the second term was $\frac{1}{3}(\frac{1}{2}x^3)$,
>
> I put $m = \frac{1}{2}$, and the terms arising were
>
> $$\frac{1}{2} \times \frac{\frac{1}{2}-1}{2} \text{ or } -\frac{1}{8}, \quad -\frac{1}{8} \times \frac{\frac{1}{2}-2}{3} \text{ or } +\frac{1}{16}, \quad \frac{1}{16} \times \frac{\frac{1}{2}-3}{4} \text{ or } -\frac{5}{128},$$
>
> and so to infinity.

After Newton saw that his formula worked for the known series, he simply applied the same rules to the series he was interested in and that got him this result. He never gave any proof or reason why he was allowed to do this, he just did it and it worked

> Whence I came to understand that the area of the circular segment which I wanted was
>
> $$x - \frac{\frac{1}{2}x^3}{3} - \frac{\frac{1}{8}x^5}{5} - \frac{\frac{1}{16}x^7}{7} - \frac{\frac{5}{128}x^9}{9} \textit{ etc.}$$

> And by the same reasoning the areas of the remaining curves, which were to be inserted, were likewise obtained: as also the area of the hyperbola and of the other alternate curves in this series
>
> $$(1-x^2)^{0/2},\ (1-x^2)^{1/2},\ (1-x^2)^{2/2},\ (1-x^2)^{3/2},\ \text{etc.}$$

Interestingly, whereas he set out to find the area of $(1-x^2)^{\frac{1}{2}}$, he ended up finding $(1-x^2)^{\frac{n}{2}}$ for any n, showing the power of generalisation in mathematics.

> And the same theory serves to intercalate other series, and that through intervals of two or more terms when they are absent at the same time.

Newton says his theory is completely general, as we can use the same process to intercalate the two terms missing in the series:

$$\begin{gathered}(1-x^2)^{\frac{0}{3}} = \text{known} \\ (1-x^2)^{\frac{1}{3}} = ? \\ (1-x^2)^{\frac{2}{3}} = ? \\ (1-x^2)^{\frac{3}{3}} = \text{known}\end{gathered} \tag{39}$$

> This was my first entry upon these studies, and it had certainly escaped my memory, had I not a few weeks ago cast my eye back on some notes.

These may well be the notes we covered in Section 2

> But when I had learnt this, I immediately began to consider that the terms
>
> $$(1-x^2)^{0/2},\ (1-x^2)^{2/2},\ (1-x^2)^{4/2},\ (1-x^2)^{6/2},\ etc.$$

> that is to say,
>
> $$1,\ 1-x^2,\ 1-2x^2+x^4,\ 1-3x^2+3x^4-x^6,\ etc.$$
>
> could be interpolated in the same way as the areas generated by them:

Here Newton goes further than the reasoning we described in Section 2, as he drops integration in his process entirely.

> and that nothing else was required for this purpose but to omit the denominators $1, 3, 5, 7$, etc., which are in the terms expressing the areas;

The denominators are the consequence of integrating monomials of increasingly higher order. Why didn't Newton stumble across his method by trying to find power series for $(1-x^2)(\frac{1}{2})$ directly? It is probably that this wasn't an interesting question. The circle was simple to understand as it has a simple definition. Creating a power series for its own sake is a fairly esoteric problem. But finding the area of a circle segment had occupied mathematicians for millennia, so it makes sense that the context of discovery of the binomial theorem involved the area under a circle segment.

> this means that the coefficients of the terms of the quantity to be intercalated $(1-x^2)^{\frac{1}{2}}$, or $(1-x^2)^{\frac{3}{2}}$, or in general $(1-x^2)^m$, arise by the continued multiplication of the terms of this series
>
> $$m \times \frac{m-1}{2} \times \frac{m-2}{3} \times \frac{m-3}{4}\ etc.$$

> so that (for example)
>
> $$(1-x^2)^{\frac{1}{2}} \text{ was the value of } 1-\frac{1}{2}x^2-\frac{1}{8}x^4-\frac{1}{16}x^6\ etc.,$$
>
> $$\text{and } (1-x^2)^{\frac{3}{2}} \text{ was the value of } 1-\frac{3}{2}x^2+\frac{3}{8}x^4+\frac{1}{16}x^6\ etc.,$$
>
> $$(1-x^2)^{\frac{1}{3}} \text{ was the value of } 1-\frac{1}{3}x^2-\frac{1}{9}x^4-\frac{5}{81}x^6\ etc.$$

So now Newton has generalized his formula to not only find the area below certain functions but also find a way to expand the functions themselves.

> So then the general reduction of radicals into infinite series by that rule, which I laid down at the beginning of my earlier letter became known to me, and that before I was acquainted with the extraction of roots.

He then generalized that to not only this specific type of function, but any function of the form $(P + PQ)^{\frac{m}{n}}$, which we now know as the binomial theorem. It is interesting that when he discovered the extraction of roots algorithm, Newton chose to give it a higher priority in the relation between mathematical ideas. In the context of justification, algorithms with infinite decimal analogies still carried more weight.

> But once this was known, that other could not long remain hidden from me.

A crucial step was of course to check that the answer makes sense:

> For in order to test these processes, I multiplied
> $$1 - \frac{1}{2}x^2 - \frac{1}{8}x^4 - \frac{1}{16}x^6, etc.$$
> into itself; and it became $1 - x^2$, the remaining terms vanishing by the continuation of the series to infinity.

To check his work, he multiplied the infinite series he found for $(1 - x^2)^{\frac{1}{2}}$ with itself, which should give $1 - x^2$, and it did. Thus he had convinced himself that it worked. He then also checked it for $(1 - x^2)^{\frac{1}{3}}$, and here it again also worked.

> And even so $1 - \frac{1}{3}x^2 - \frac{1}{9}x^4 - \frac{5}{81}x^6$, etc. multiplied twice into itself also produced $1 - x^2$.

> And as this was not only sure proof of these conclusions so too it guided me to try whether, conversely, these series, which it thus affirmed to be roots of the quantity $1 - x^2$, might not be extracted out of it in an arithmetical manner.

So after multiplying his series by itself to check that it indeed makes $(1 - x^2)$, Newton is prompted to find out whether you can also apply the same 'root extraction algorithm' that is commonly applied to decimal numbers, to algebraic quantities like $(1 - x^2)^{\frac{1}{2}}$:

> And the matter turned out well. This was the form of the working in square roots.
>
> $$\begin{array}{llllll} 1 & -x^2(1 & -\frac{1}{2}x^2 & -\frac{1}{8}x^4 & -\frac{1}{16}x^6 & \text{, etc.} \\ 1 & & & & & \\ \hline 0 & -x^2 & & & & \\ & -x^2 & +\frac{1}{4}x^4 & & & \\ \hline & & -\frac{1}{4}x^4 & & & \\ & & -\frac{1}{4}x^4 & +\frac{1}{8}x^6 & +\frac{1}{64}x^8 & \\ \hline & & 0 & -\frac{1}{8}x^6 & -\frac{1}{64}x^8 & \end{array}$$

We addressed this algorithm in Section 3.

> After getting this clear I have quite given up the interpolation of series, and have made use of these operations only, as giving more natural foundations.

Again, we stress that Newton is operating in the context of justification - in his day-to-day life he is more likely to have used the binomial theorem, as it "greatly simplifies"[29] root extractions.

Newton then says that just like we can carry out the root extraction algorithm for decimals as we normally apply it to numbers, we can do the same ("nor was there any secret") with

[29] Newton, 'Epistola Prior (Letter to Gottfried Wilhelm Leibniz via Henry Oldenburg, June 13, 1676)'.

the reduction by division algorithm. It is "an easier affair", so bringing the binomial theorem when doing things like $(1+x)^{-1}$ is not as useful to him, but it's possible:

> Nor was there any secret about reduction by division, an easier affair in any case.

Bibliography

Dennis, David, and Jere Confrey. 'The Creation of Continuous Exponents: A Study of the Methods and Epistemology of John Wallis'. In *CBMS Issues in Mathematics Education*, vol. 6, *CBMS Issues in Mathematics Education*, edited by James Kaput, Alan Schoenfeld, Ed Dubinsky, and Thomas Dick. American Mathematical Society, 1996. https://doi.org/10.1090/cbmath/006/02.

June Barrow-Green, Jeremy Gray, and Robin Wilson. *The History of Mathematics: A Source-Based Approach Vol 2*. 2020.

Mercator, Nicolaus. *Logarithmotechnia: Sive, Methodus Construendi Logarithmos Nova, Accurata, Et Facilis*. Guilielmi Godbid, 1668.

Newton, Isaac. 'Epistola Posterior (Letter to Gottfried Wilhelm Leibniz via Henry Oldenburg, October 24, 1676)'. In *The Correspondence of Isaac Newton*, 2 (1676–1687), *The Correspondence of Isaac Newton*, edited by Herbert W. Turnbull. Cambridge University Press, 1960.

Newton, Isaac. 'Epistola Prior (Letter to Gottfried Wilhelm Leibniz via Henry Oldenburg, June 13, 1676)'. In *The Correspondence of Isaac Newton*, 2 (1676–1687), *The Correspondence of Isaac Newton*, edited by Herbert W. Turnbull. Cambridge University Press, 1960.

Newton, Isaac. 'Mathematical Papers (Cambridge University Library, MS Add. 3958:72)'. England, 1665. https://cudl.lib.cam.ac.uk/view/MS-ADD-03958/139.

Newton, Isaac. *De Analysi Per Æquationes Numero Terminorum Infinitas*. Ex officina Pearsoniana, 1711.

Stedall, Jacqueline. *Mathematics Emerging: A Sourcebook 1540–1900*. Oxford University Press, 2008. https://nx89456.your-storageshare.de/s/ArjNPeHYjkom6Zt.

Strogatz, Steven. 'Strogatz_20newton_20series'. 2019. https://www.stevenstrogatz.com/essays/2019/11/19/appendix-2-unpublished-from-infinite-powers.

Strogatz, Steven. *Infinite Powers: How Calculus Reveals the Secrets of the Universe*. First Mariner books edition. Mariner Books ; Houghton Mifflin Harcourt, 2020.

Wallis, John, and Jacqueline A. Stedall. *The Arithmetic of Infinitesimals*. Sources and Studies in the History of Mathematics and Physical Sciences. Springer, 2004.

Whiteside, D. T. 'Newton's Discovery of the General Binomial Theorem'. *The Mathematical Gazette* 45, no. 353 (1961): 175–80. https://doi.org/10.2307/3612767.