Draft

P1 – Group 7

Line-by-Line Commentary

Discovering the Binomial Theorem

Victor Elgersma & Vera Belde

Submitted 10 March 2026

Updated 23 April 2026

Contents

[1] https://uu.brightspace.com/d2l/le/lessons/44275/topics/498036, page 191

[2] from Newton, 'De analysi', 1669, as published in Analysis per quantitatum series, fluxiones, ac differentias, 1711, 6–7, available on https://uu.brightspace.com/d2l/le/lessons/44275/topic, page 194

1. Unpublished Manuscript CUL Add. MS 3958.3, f. 72.[3]

In this page from a young Newton's handwritten notes, we find the integrals of $\frac{1}{x}$ and $\sqrt{1-x^2}$ as power series, which Newton 'intuits' by extrapolating the pattern created by a broader class of easier integrals. In his second letter to Leibniz, he explains that this is how he first stumbled upon these series, but notes that polynomial division and the root extraction algorithm provide a firmer foundation.

What follows is not a proof in a sense that a modern mathematician would recognize[4], but we do think it is pretty convincing.

We first turn to the hyperbola.

1.1. Integral of the Hyperbola

> 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).

This boils down to "draw the curves $y = \frac{1}{x'}, y = 1, y = x', y = x'^2$ and $y = x'^3$ in the $x' - y$ plane. For each of these curves he notes the values at $x' = x + 1$

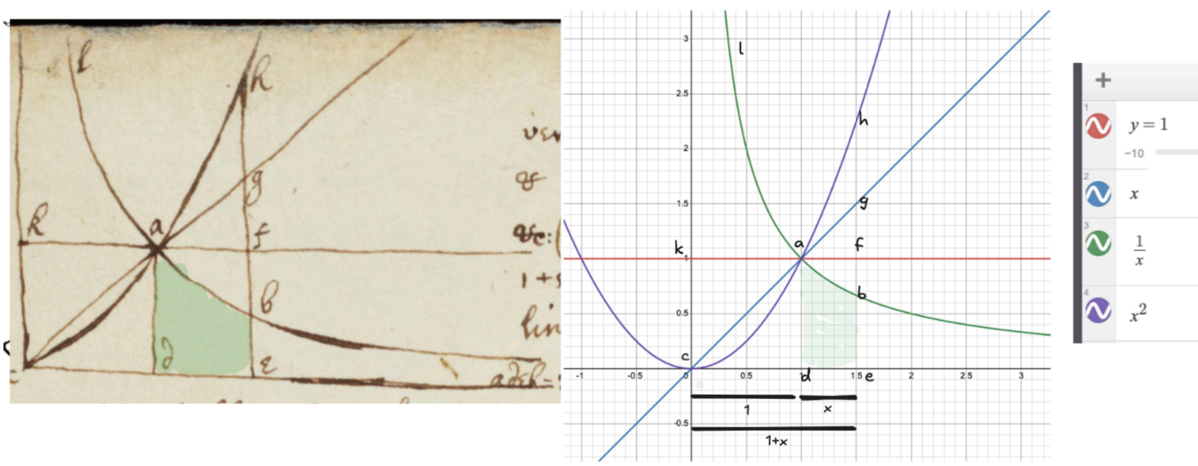


Figure 1: Sketch to show the curves Newton is considering to find the integral of $\frac{1}{x'}$ from 1 to 1+x, on the right is a facelift done with Desmos

[3] https://uu.brightspace.com/d2l/le/lessons/44275/topics/498036, page 191

[4] For one, Newton manipulates infinite series as if they were finite, without first checking that the series converges

We use x' and not x because x is reserved for the distance between the points d ($x' = 1$) and e ($x' = 1 + x$). This may seem strange but is necessary for Pascal's triangle coefficients to appear in the integrals that follow.

> 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.

In other words, we calculate the following integrals:

$$\int_1^{1+x} \frac{1}{x'}\,\mathrm{d}x' = \star$$

$$\int_1^{1+x} \mathrm{d}x' = x\ |_1^{1+x} = (1+x) - 1 = \mathbf{1}x$$

$$\int_1^{1+x} x'\,\mathrm{d}x' = \frac{x'^2}{2}\ |_1^{1+x} = \frac{(1+x)^2 - 1}{2} = \mathbf{1}x + \mathbf{1}\frac{x^2}{2}$$

$$\int_1^{1+x} x'^2\,\mathrm{d}x' = \frac{x'^3}{3}\ |_1^{1+x} = \frac{(1+x)^3 - 1}{3} = \mathbf{1}x + \mathbf{2}\frac{x^2}{2} + \mathbf{1}\frac{x^3}{3}$$

$$\int_1^{1+x} x'^3\,\mathrm{d}x' = \frac{x'^4}{4}|_1^{1+x} = \frac{(1+x)^4 - 1}{4} = \mathbf{1}x + \mathbf{3}\frac{x^2}{2} + \mathbf{3}\frac{x^3}{3} + \mathbf{1}\frac{x^4}{4}$$

$$\int_1^{1+x} x'^4 dx' = \frac{x'^5}{5}\ |_1^{1+x} = \frac{(1+x)^5 - 1}{5} = \mathbf{1}x + \mathbf{4}\frac{x^2}{2} + \mathbf{6}\frac{x^3}{3} + \mathbf{4}\frac{x^4}{4} + \mathbf{1}\frac{x^5}{5}$$

$$\int_1^{1+x} x'^5 dx' = \frac{x'^5}{5}\ |_1^{1+x} = \frac{(1+x)^6 - 1}{6} = \mathbf{1}x + \mathbf{5}\frac{x^2}{2} + \mathbf{10}\frac{x^3}{3} + \mathbf{10}\frac{x^4}{4} + \mathbf{5}\frac{x^5}{5} + \mathbf{1}\frac{x^6}{6}$$

We have emphasized the coefficients of Pascal's triangle in **bold**. Newton then draws a table of the coefficients of the integrals he has, which allows him to extrapolate the pattern to establish his best guess for $\star$:

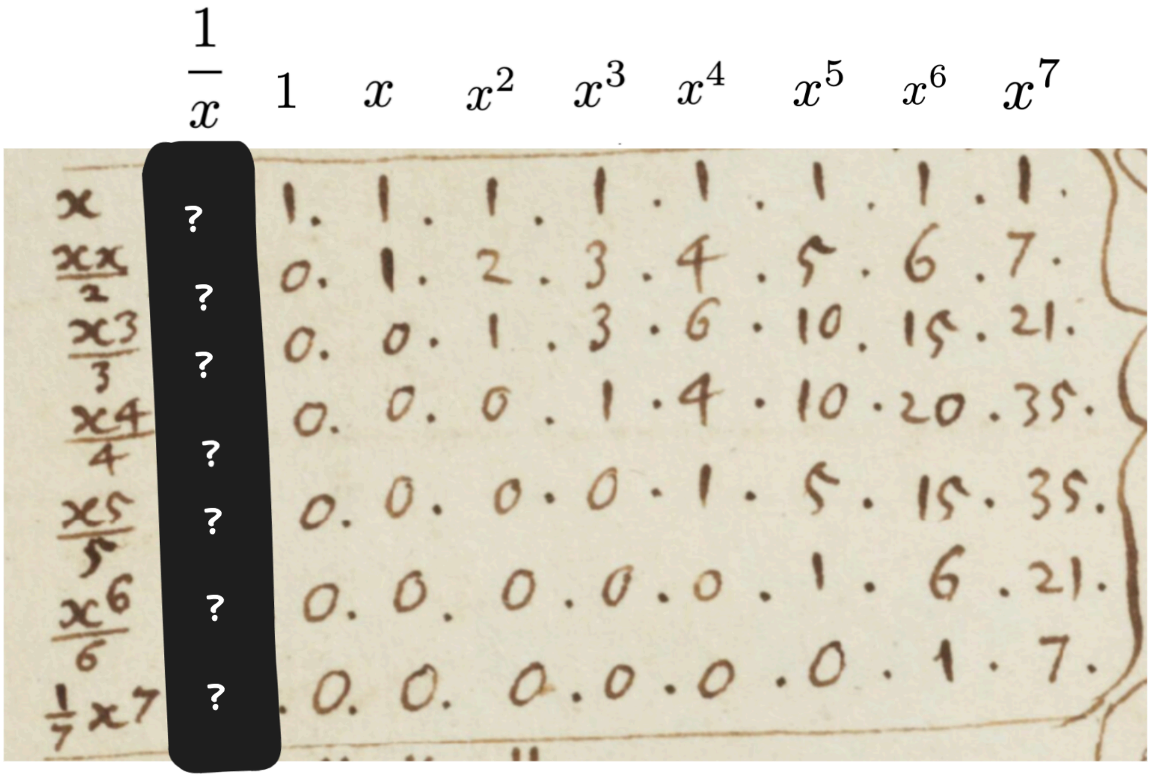


Figure 2: Each column represents the integral of a polynomial, where the coefficients are the same as the ones above in **bold**. Rows are increasingly higher terms in the expansions of the integrals. The integral of $\frac{1}{x}$ is the punchline of the 'proof' and we have redacted it at this stage. For the unredacted version see Figure 3.

> ***As in this table. In which the first area is also inserted.***

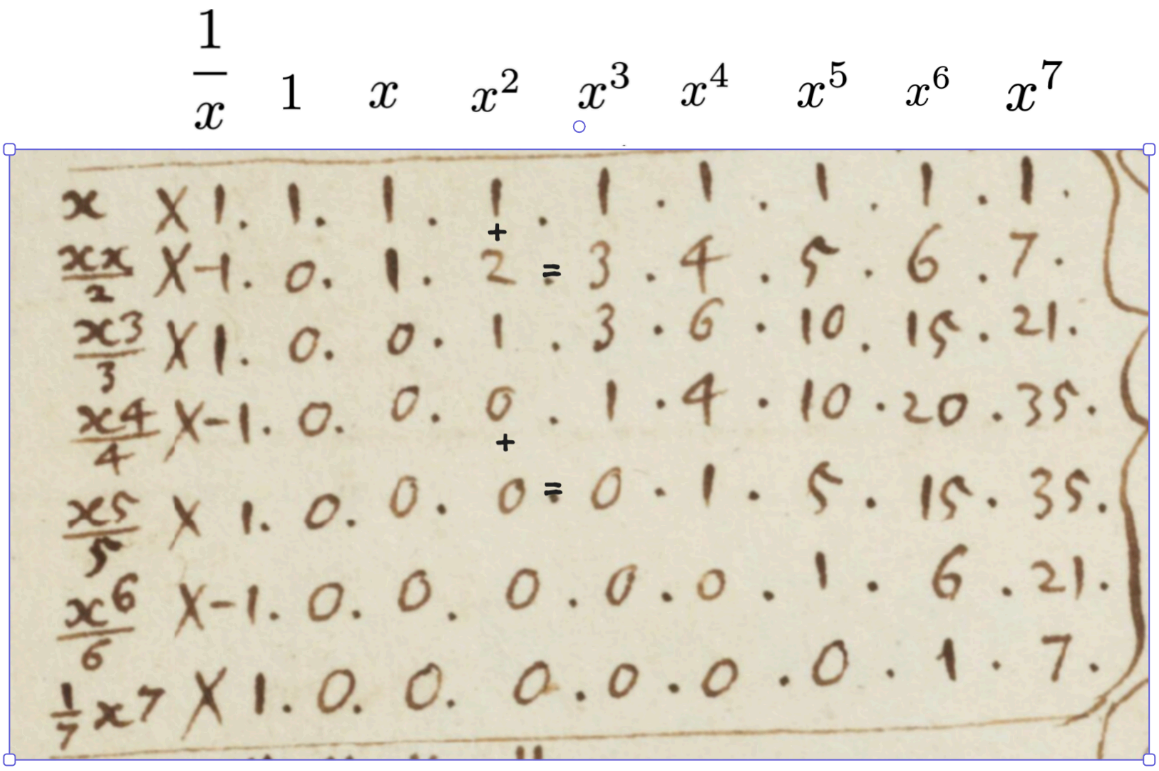


Figure 3: Same as Figure 2 but with the coefficients for the integral of $\frac{1}{x}$ included. I have added some $+$ and $=$ signs to illustrate Newton's statement that *"The sum of any figure and the figure above it is equal to the figure next to it"*

> 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.

By writing the coefficients in a table and viewing Pascal's triangle as a square, this allows us to infer a 'missing' sequence which in this case coincides with the integral of 1/x.

Newton doesn't explicitly state his reasoning for the coefficients of the integral of $\frac{1}{x}$, but we can guess the argument goes something like this: First, the top-left element is probably 1, because all the topmost elements are 1. Second, work out all the subsequent numbers by the rule that "The sum of any figure and the figure above it is equal to the figure following it". This leads Newton to his punchline:

> 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.

Newton is guarded in his conclusions. “May appear” is not the same as “has been shown”.[5] And indeed, all he has done is provide some numerical evidence that *if* there is a power series expansion for $\frac{1}{x}$, then this particular expansion would respect the pattern.

Next, Newton turns his attention to integrating a half-circle between arbitrary points on the x-axis.

1.2. Integral of Circle

While Wallis[6] had expressed π as a power series, Newton generalized his result by finding the integral of a semi-circle between any two points on the x-axis.

> 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.

Like with the hyperbola, Newton begins his study of the area of a circle by drawing an array of curves with the form $(1-x^2)^m$, for both integer and half-integer m.

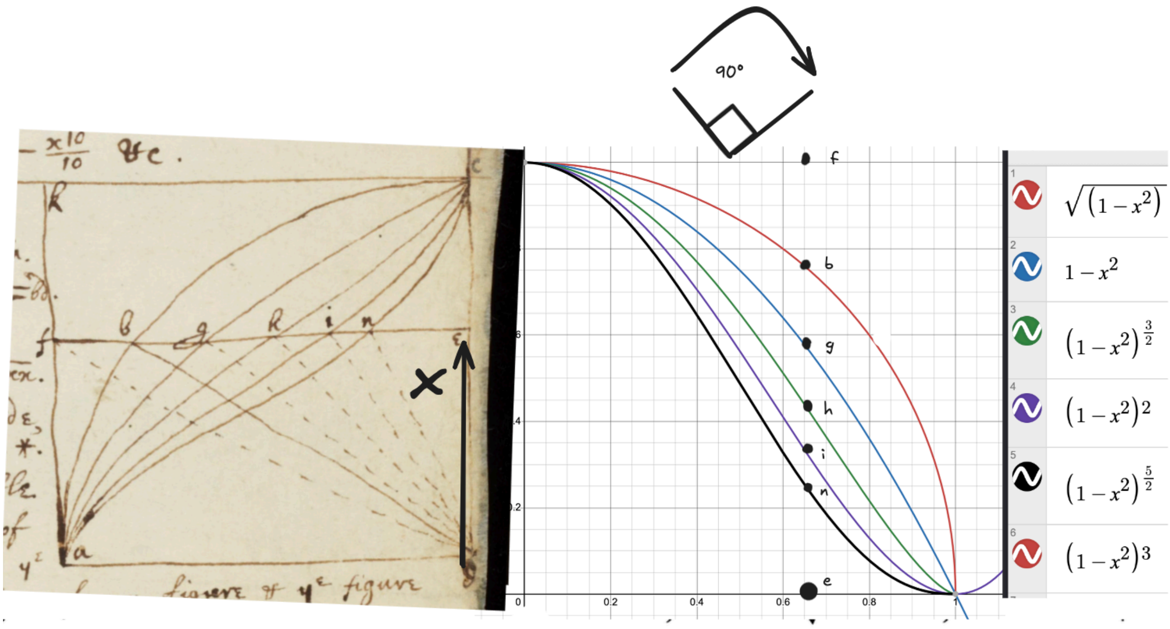


Figure 4: Left: Newton’s original sketch. Right: We rotated Newton’s sketch 90° clockwise to make sense of it. In modern notation, he is looking for patterns in the integrals of the form $\int_0^x (1-x^2)^m dx$ for $m = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}$ etc.

Newton then proceeds to integrate the easy ones:

[5] This observation was made by Niccolò Guicciardini, Università degli Studi di Milano in his video exposition of this manuscript, available at: https://cudl.lib.cam.ac.uk/view/MS-ADD-03958/139

[6] As claimed in Strogatz, https://uu.brightspace.com/d2l/le/lessons/44275/topics/498037

> 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:

In other words, he has just done these integrals:

$$\int_0^x (1-x^2)^0 \, \mathrm{d}x = \mathbf{1}x$$

$$\int_0^x (1-x^2)^{\frac{1}{2}} \, \mathrm{d}x = \star$$

$$\int_0^x (1-x^2) \, \mathrm{d}x = \mathbf{1}x + \mathbf{1}\left(-\frac{x^3}{3}\right)$$

$$\int_0^x (1-x^2)^{\frac{3}{2}} \, \mathrm{d}x = \star$$

$$\int_0^x (1-x^2)^2 \, \mathrm{d}x = \int_0^x (1-2x^2+x^4) \, \mathrm{d}x = \mathbf{1}x + \mathbf{2}\left(-\frac{x^3}{3}\right) + \mathbf{1}\frac{x^5}{5}$$

$$\int_0^x (1-x^2)^{\frac{5}{2}} \, \mathrm{d}x = \star$$

$$\int_0^x (1-x^2)^3 \, \mathrm{d}x = \int_0^x (1-3x^2+3x^4-x^6) \, \mathrm{d}x = \mathbf{1}x + \mathbf{3}\left(-\frac{x^3}{3}\right) + \mathbf{3}\frac{x^5}{5} + \mathbf{1}\left(-\frac{x^7}{7}\right)$$

$$\int_0^x (1-x^2)^{\frac{7}{2}} \, \mathrm{d}x = \star$$

$$\int_0^x (1-x^2)^4 \, \mathrm{d}x = \int_0^x (1-4x^2+6x^4-4x^6+x^8) \, \mathrm{d}x = \mathbf{1}x + \mathbf{4}\left(-\frac{x^3}{3}\right) + \mathbf{6}\frac{x^5}{5} + \mathbf{4}\left(-\frac{x^7}{7}\right) + \mathbf{1}\frac{x^9}{9}$$

Again, we have arranged things such that Pascal's triangle coefficients are now in **bold**. Note also the alternating signs between terms.

> as in this table following in which the indeterminate terms are inserted.

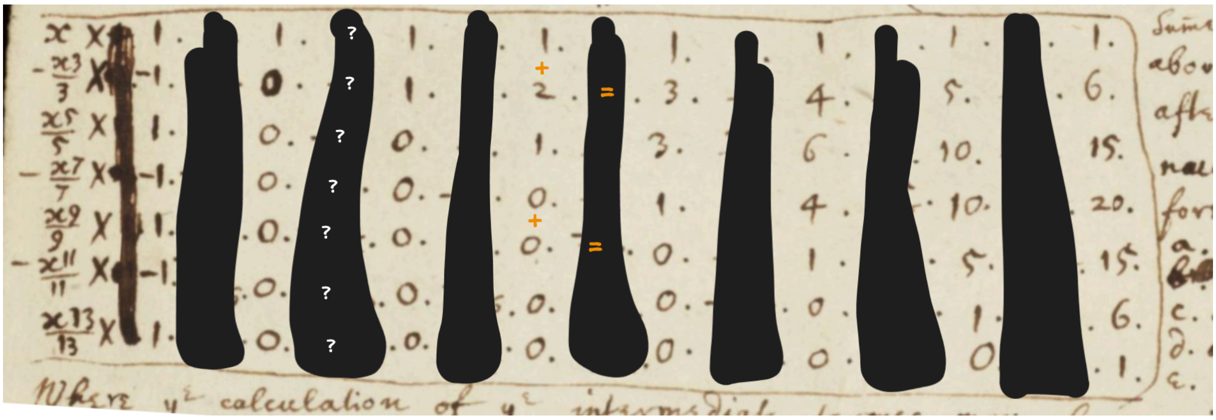


Figure 5: Each column represents the integral of a polynomial of the form $(1-x^2)^m$ for $m = -1, \frac{1}{2}, 0, \frac{1}{2}, 1, \frac{3}{2}$ except we have redacted the values $m = -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}$ etc. The fourth column (redacted) contains the case of the circle $m = \frac{1}{2}$. Newton does not draw the graphs for $m = -1, -\frac{1}{2}$ but he does does treat them in this table. Each row represents a term in the series expansion for the aforementioned column.

Like in his treatment of the hyperbola, Newton spots Pascal's triangle:

> 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.

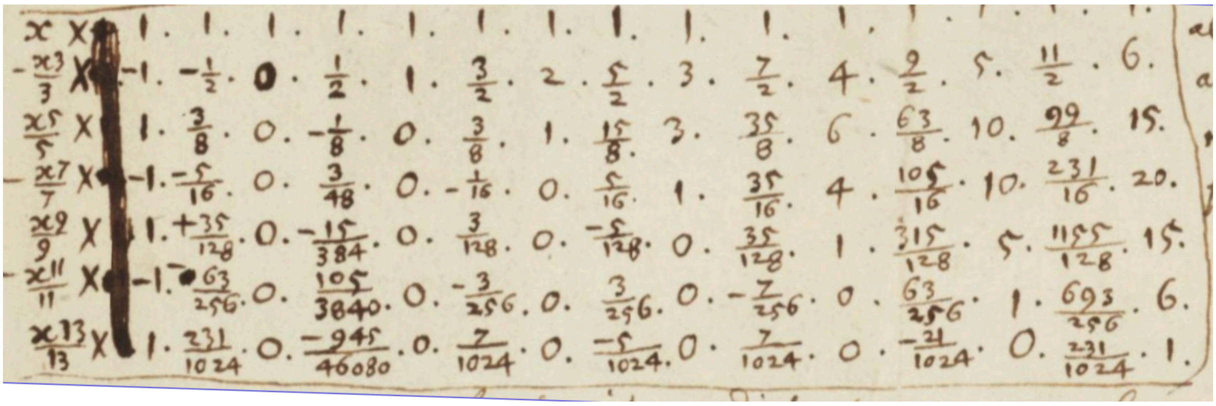
Newton then creates the following small finite difference table[7]

> 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.

[7]We follow Stedall in calling this a 'finite difference table'; although we do not quite understand how this table leads him to conclude the form of the binomial coefficients. We do see that substituting a=1 gives us back the coefficients of Pascal's triangle.



x	×	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$-\frac{x^3}{3}$	×	−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
$\frac{x^5}{5}$	×	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
$-\frac{x^7}{7}$	×	−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
$\frac{x^9}{9}$	×	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
$-\frac{x^{11}}{11}$	×	−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
$\frac{x^{13}}{13}$	×	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 6: The unredacted table which includes the intermediate terms

We do not see how the calculation of the intermediate terms 'may be easily performed' from the finite difference table unless we take into account Newton's second letter to Leibniz, in which he explains that he obtained a general formula for the binomial coefficients for integer m, and then plugged in $m = \frac{1}{2}$ to get the value for the circle.

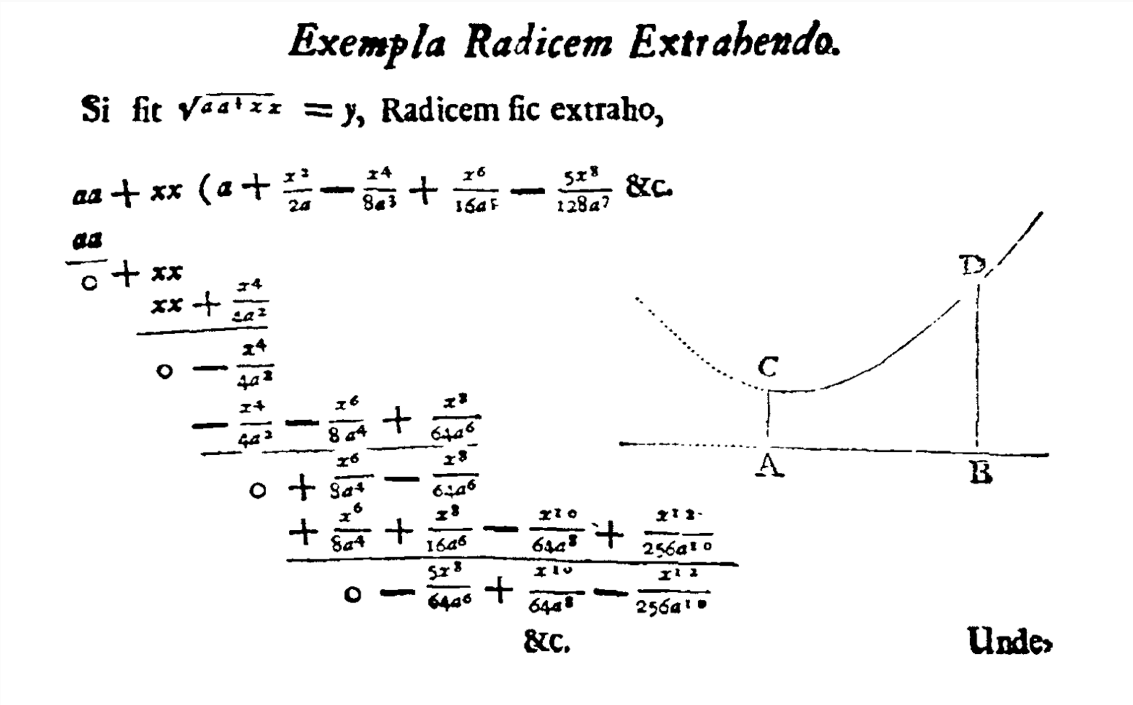
The logic goes something like this: First, the coefficient of x in the expansion of $\sqrt{1 - x^2}$ must be 1, since the first row consists entirely of 1s. So $a = 1$. The coefficient of $-\frac{x^3}{3}$ must be $\frac{1}{2}$ because the second row is an arithmetic series $-1, 0, 1, 2, 3, 4$. So $b = \frac{1}{2}$. Next, we spot the following formula for generating successive terms of Pascal's triangle. If the second row is m, then the third term is $m \times \frac{m-1}{2}$, the fourth term is $m \times \frac{m-1}{2} \times \frac{m-2}{3}$, the fifth term is $m \times \frac{m-1}{2} \times \frac{m-2}{3} \times \frac{m-3}{4}$ and so on. For a fuller explanation of this please see our section on Newton's Epistola Posterior.

This is what what we would today recognize as the binomial coefficient, i.e the statement that the k^{th} term of the binomial expansion of $\left(1 - x^2\right)^m$ is: $\frac{m!}{(m-k)!k!}(-1)^k \frac{x^{2k+1}}{2k+1}$.

2. Root Extraction: A Series from Newton's 'De analysi'[8]

From Newton, 'De analysi', 1669, as published in *Analysis per quantitatum series fluxiones, ac differentias*, 1711, 6-7.

In the following extract, Newton shows another way of obtaining infinite series - by adapting an algorithm that was commonly used at the time for getting the square roots of numbers.



Exempla Radicem Extrahendo.

Si sit $\sqrt{aa+xx}=y$, Radicem sic extraho,

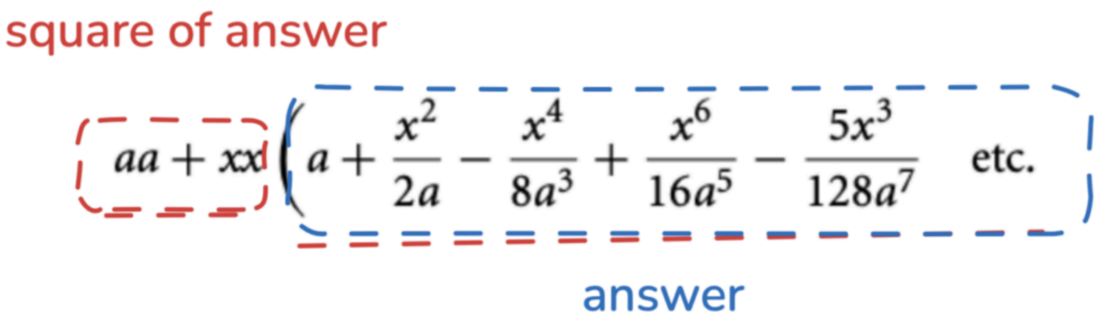
$$aa+xx\ \left(a+\frac{x^2}{2a}-\frac{x^4}{8a^3}+\frac{x^6}{16a^5}-\frac{5x^8}{128a^7}\ \text{\&c.}\right.$$

$$\begin{array}{l} aa \\ \hline 0+xx \\ \quad xx+\frac{x^4}{4a^2} \\ \hline \quad 0-\frac{x^4}{4a^2} \\ \quad -\frac{x^4}{4a^2}-\frac{x^6}{8a^4}+\frac{x^8}{64a^6} \\ \hline \quad 0+\frac{x^6}{8a^4}-\frac{x^8}{64a^6} \\ \quad +\frac{x^6}{8a^4}+\frac{x^8}{16a^6}-\frac{x^{10}}{64a^8}+\frac{x^{12}}{256a^{10}} \\ \hline \quad 0-\frac{5x^8}{64a^6}+\frac{x^{10}}{64a^8}-\frac{x^{12}}{256a^{10}} \end{array}$$

&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:



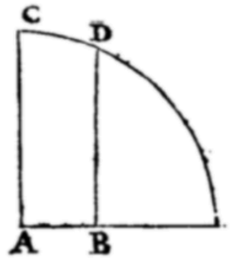
The algorithm goes as follows: We start with guessing a number whose square will equal a^2. This is obviously a. Then we take our guess (a), square it (a^2), and take it away from our target answer $a^2 + x^2$. The remainder is x^2. We take the remainder, divide it by $2a$, and add this value to our 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,

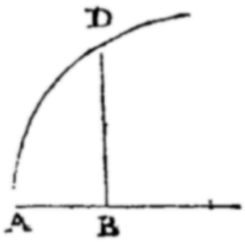
[8]from Newton, 'De analysi', 1669, as published in Analysis per quantitatum series, fluxiones, ac differentias, 1711, 6–7, available on https://uu.brightspace.com/d2l/le/lessons/44275/topic, page 194

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.

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

> 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.





“Et haec est Quadratura Hyperbola” - thus we have squared the hyperbola.

3. Epistola Prior (1676)

This is an english translation of the letter sent from Newton to Leibniz (via Oldenburg). The original is in Latin. 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}.

> 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;

> and quantities by extraction of the roots,

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

Although the ‘algorithmic approach’ to generating infinite series, is, according to Newton, the “most foundational” of the methods, it is not the first he stumbled across. This also explains his enthusiasm for having multiple ways for arriving at the same truth: sometimes the most foundational method is not the first to present itself.

> 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.

It is interesting that he says "extraction of roots" are much shortened by the binomial theorem, as opposed to saying "both reduction by division and extraction of roots" are much shortened by this theorem, since the binomial theorem does apply to negative powers, and hence can replace polynomial division problems like $\frac{1}{1+x}$. Perhaps the root extraction algorithm is much more laborious than polynomial division, which explains why the former justifies a short-cut, whereas the latter doesn't.

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 $\frac{1}{a}$:

> 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, 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.

> 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.

Today, we rarely see the binomial theorem written like $(x+y)^n = \left(x + x\left(\frac{y}{x}\right)\right)^n$ in mathematics textbooks, although we do still find this step used a lot in physics contexts and small-parameter approximations: students tend to memorize the first 1 or 2 terms of the binomial series, namely $(1+d)^n = 1 + nd + \frac{n(n-1)}{2}d^2$. Then, getting an equation like $(x+y)^n$ into the form $x^n\left(1+\frac{y}{x}\right)^n$ for $\frac{y}{x} \ll 1$ is very useful indeed, as it allows us to quickly expand it into the first few terms.

> 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. If Newton had more space, it may have been clearer if he wrote his theorem as:

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

where:

$$A = P^{\frac{m}{n}}$$
$$B = \frac{m}{n}AQ$$
$$C = \frac{m-n}{2n}BQ$$
$$D = \frac{m-2n}{3n}CQ$$

Which makes the recursive nature immediately obvious.

3.1. Showing the equivalence between the modern $\sum \binom{r}{k} x^{r-k} y^k$ and Newton's $A + B + C + D + E + \ldots$

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:

$$(x+y)^r = \sum_{k=0}^{r \text{ or } \infty} \binom{r}{k} x^{r-k} y^k \qquad \text{where:} \quad \binom{r}{k} = \frac{r!}{k!(r-k)!}$$

Where the sum is finite for integer p and infinite for fractional (i.e non-integer but rational) p, and:

Let us now show the equivalence of the modern statement of the Binomial Theorem and Newton's recursive one.

We start by writing out the first few terms:

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

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

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

now we note that r is actually the fraction $\frac{m}{n}$:

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

We can simplify some of the fractions by multiplying through by n:

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

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

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

Now, $B = \frac{m}{n} AQ$, so we can say there's a B hiding in the C and D terms:

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

Similarly, there a C hiding in the D term:

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

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.

3.2. 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:

$$m = 1$$

$$n = 2$$

$$P + PQ = 10$$

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:

$$Q = \frac{1}{9}$$

$$P = 9$$

Then

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

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

$$(P+PQ)^{\frac{m}{n}} =$$

$$3 + \left(\frac{1}{2}\right)A\left(\frac{1}{9}\right) + \frac{1-2}{(2)(2)}\left(B\left(\frac{1}{9}\right)\right) + \frac{1-2(2)}{(3)(2)}\left(C\left(\frac{1}{9}\right)\right) + \frac{1-3(2)}{(4)(2)}\left(D\left(\frac{1}{9}\right)\right)$$

$$A = 3$$

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

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

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

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

Indeed

$$(A+B+C+D+E)^2 = 9.99999188742702$$

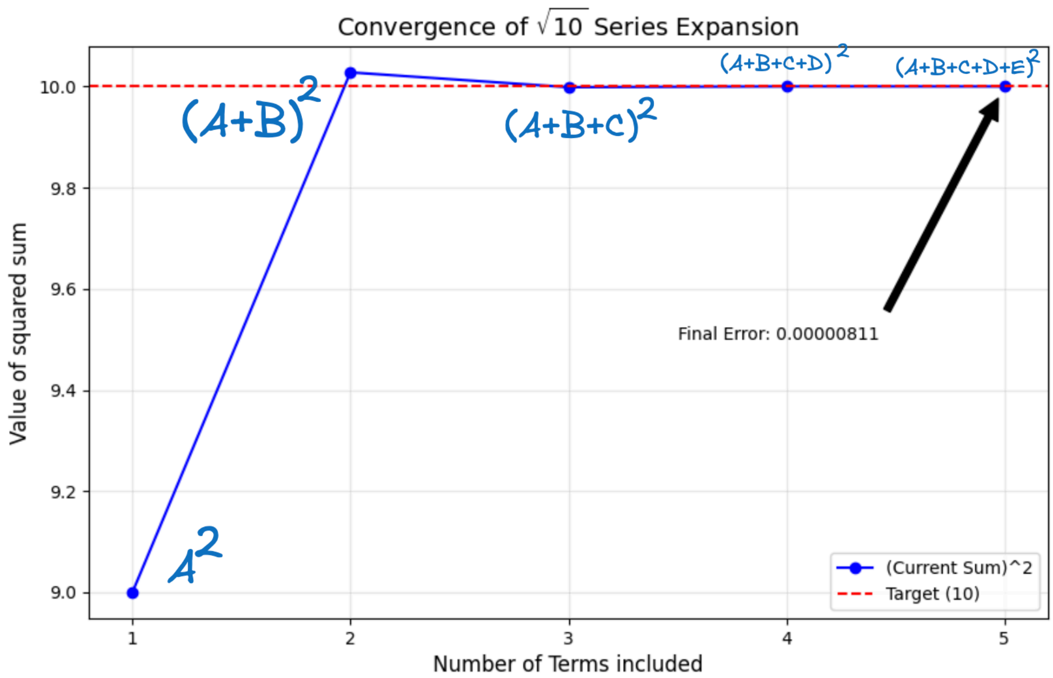


Figure 8: Graph to show rapid conversion of the binomial expansion of square root of 10

Newton next gives us some algebraic examples, which also show the application of his new notation for roots. The first example is exactly of the form $(P + PQ)^{\frac{m}{n}}$

> *Example 1*
>
> $\sqrt{(c^2 + x^2)}$ or $(c^2 + x^2)^{1/2} = c + \frac{x^2}{2c} - \frac{x^4}{8c^3} + \frac{x^6}{16c^5} - \frac{5x^8}{128c^7} + \frac{7x^{10}}{256[c]^9} +$ *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 second example has a typo as $Q = \frac{c^4x - x^5}{c^5}$, in the first case, not $\frac{(c^4x-x)^5}{c^5}$.

And indeed, whether to identify P with c^5 or with $-x^5$ depends on which has the greater magnitude.

> *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.

4. Epistola Posterior (1676)

This is the introduction of the second letter Newton sent to Leibniz (via Oldenburg) about the infinite series he had found. In this letter, Newton appears explains how he found the binomial series as stated in his first letter.

> 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.

> 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.

Again there appears to be a typo in the translation as the first term ought to be x not x^3

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.

> 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:

$$\int (1-x^2)^{\frac{0}{2}} dx = \int 1 dx = x$$

$$\int (1-x^2)^{\frac{1}{2}} dx$$

$$\int (1-x^2)^{\frac{2}{2}} dx = \int (1-x^2) dx$$

$$\int (1-x^2)^{\frac{3}{2}} dx$$

$$\int (1-x^2)^{\frac{4}{2}} dx = \int (1-x^2)^2 dx = \int (1-2x^2+x^4) dx$$

$$\int (1-x^2)^{\frac{5}{2}} dx$$

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

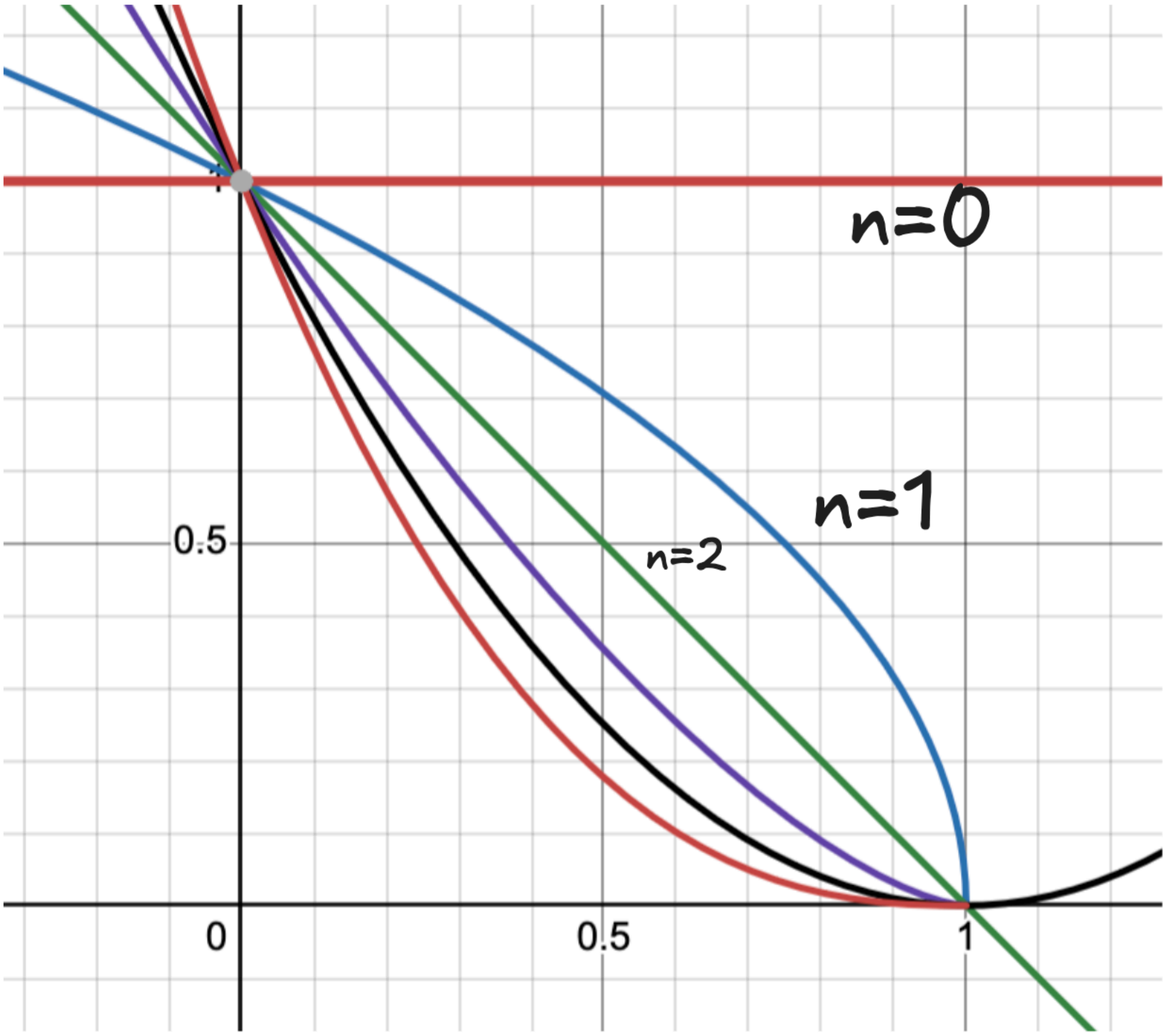


Figure 9: $(1-x^2)^{\frac{n}{2}}$ plotted

> 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.

> $1, 2, 1$; 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,
>
> $$\frac{m-0}{1} \times \frac{m-1}{2} \times \frac{m-2}{3} \times \frac{m-3}{4} \times \frac{m-4}{5} \times etc.$$

> 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.

> 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.

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

> 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
>
> $$(1-x^2)^{0/2},\ (1-x^2)^{1/2},\ (1-x^2)^{2/2},\ (1-x^2)^{3/2},\ \text{etc.}$$

> 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.

> 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 1

> 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:

> 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;

> 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 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.

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

> 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, but to be more rigorous, he 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.
>
> $$\begin{array}{llllll} 1 & -x^2(1 & -\frac{1}{2}x^2 & -\frac{1}{8}x^4 & -\frac{1}{16}x^6 & \text{, etc.} \\ 1 & & & & & \\ \hline 0 & -x^2 & & & & \\ & -x^2 & +\frac{1}{4}x^4 & & & \\ \hline & & -\frac{1}{4}x^4 & & & \\ & & -\frac{1}{4}x^4 & +\frac{1}{8}x^6 & +\frac{1}{64}x^8 & \\ \hline & & 0 & -\frac{1}{8}x^6 & -\frac{1}{64}x^8 & \end{array}$$

> 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.

It seems hard to believe that Newton would choose to use those operations as opposed to simply using the binomial theorem, especially since the binomial theorem would work for unusual integer powers like $\frac{3}{4}$, whereas we know of no straightforward root extraction algorithm for arbitrary roots. So we should read this as “I use these operations, as opposed to the interpolation of series, to justify my use of the binomial theorem”, especially since in the *Epistola Prior* (Section 3) he writes that “the result is greatly simplified by the [binomial theorem]”

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 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