General binomial expansion

Shrivathsa Pandelu

October 9, 2020

Contents

1 Introduction

One day (19th September 2019), as I was going through some things from Lang's Complex analysis, I came across the general binomial expansion formula. He stated the theorem and found the radius of convergence. He wrote

$$(1+z)^\alpha = \sum_{n\geq 0} \binom{\alpha}{n} z^n$$

where $\binom{\alpha}{n}$ is the general binimal coefficient. However, I couldn't find anywhere in the book where he shows that this expansion is indeed $(1+z)^\alpha$, that is whether raising this to α^{-1} would give back $1+z$. So, I tried to prove it on my own, which lead me to a rabit hole that has resulted in this essay.

Whenever I write $[a,b]$ I mean a proper closed interval with $-\infty < a < b < \infty$.

2 Preliminaries

I am not going to construct the real line, I will take it for granted. All I need is that there is such a set called the real line, which contains the rationals and is ordered etc. I will also need the completeness axiom. This says that given any set bounded above, there is a supremum, the least upper bound.

A sequence in $\mathbb{R}$ is a function $f\colon \mathbb{N} \to \mathbb{R}$. The limit of a sequence, if it exists, is an element $x \in \mathbb{R}$ satisfying

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } |x - f(n)| < \epsilon \forall n \geq N.$$

A limit may not always exist but when it does, it is unique. It is common to write f_n for $f(n)$.

Given two sequences, it is possible to add, multiply and divide them, and in all these cases, the limits behave similarly, that is they are added, or multiplied, or divided. Of course, one needs to be carefull when talking about division.

Next, we discuss continuous functions. Intuitively, a function is continuous if it takes points that are close to each other to points that are close. So, we need to define what "closeness" is. In $\mathbb{R}$, this is possible by taking the distance between two points. In general, we define neighbourhoods of a point called open sets. In the context of real numbers, if $f\colon \mathbb{R} \to \mathbb{R}$ is a function, we want $f(x), f(y)$ to be close whenever x, y are close. Further, no matter how close we get to $f(x)$, we want points near x to go that close to $f(x)$, that is given any $\epsilon > 0$, we should be able to find points y close to x, such that $f(x), f(y)$ are within ϵ of $f(x)$. Mathematically, we want

$$\forall \epsilon > 0 \,\exists\, \delta > 0 \text{ such that } |x - y| < \delta \implies |f(x) - f(y)| < \epsilon.$$

In general topological spaces, we would like the inverse of open sets to be open, that is the set of points that go near $f(x)$ should contain points close to x which is another way of saying that the inverse of neighbouhoods of $f(x)$ should contain neighbourhoods of x. Note that points far from x may also land close to $f(x)$, but we are looking at points near x and forcing them to land near $f(x)$.

Note that if "no matter how close we get to $f(x)$, we want points near x to go that close to $f(x)$" doesn't hold, then the points near x aren't going to points near $f(x)$. It is tempting to think that a continuous function would to the opposite, which is to take open sets to open sets. However, this is not what we want, for consider a function f which tears an open set, something like cutting an interval into two disjoint sets. Then f takes open sets to open sets, but is not continuous. Taking open sets to open sets just makes sure that the image of x and points near x have some points nearby, these points need not be the images of points near x. To be more clear, what it does is that $f(x)$ is close to many $f(y)$, which is to say that $f(x)$ contains points near it in the image of an open set containing x, however, these y need not be near x, there is no guarantee of that.

In $\mathbb{R}$, or in any other metric space, continuity is the same as saying that if $\{x_n\}_{n\geq 0}$ is a sequence which converges to x, then $\{f(x_n)\}_{n\geq 0}$ converges to $f(x)$. Using this, we can show that adding, multiplying and dividing continuous funcitons results in continuous functions. More, generally, one can show that the operations $+, \times$ are continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$. This makes $\mathbb{R}$ what is called a topological group/ring.

Next, we have the **Intermediate value theorem** which says that if f is a continuous function on a closed interval $[a,b] \to \mathbb{R}$, then f attains all values between $f(a), f(b)$. More generally, we have that the continuous image of a connected topological space is connected.

3 Exponentiation

3.1 Rational powers

For a number $a \in \mathbb{R}_{>0}$, we define (this definition works because multiplication is associative), for a positive integer n, the nth power of a by

$$a^n = a \times \cdots \times a, n \text{ times} .$$

We then define

$$a^{-n} = \frac{1}{a} \times \cdots \times \frac{1}{a}, n \text{ times}$$

for positive n and $0^n = 0$ for all $n > 0$. We also define $a^0 = 1$. Next, it is evident that there are numbers such as $\sqrt{2}$, so we want to define the rational powers of a. Since we are confined, at the moment, to $\mathbb{R}$, we know that there is no such number which is $\sqrt{-1}$. So, it is unclear as to which $a^{\frac{1}{n}}$ exist.

Firstly we observe some arithmetic properties of exponentiation. Given positive real numbers a, b, and integers n_1, n_2, by the properties of $\mathbb{R}$ (like commutativity of multiplication), it is clear that

$$\begin{aligned} a^{n_1+n_2} &= a^{n_1}a^{n_2} \\ (a^{n_1})^{n_2} &= a^{n_1n_2} \\ (ab)^{n_1} &= a^{n_1}b^{n_2} \end{aligned}$$

For any positive $a \neq 1$, if $a^n = 1$ for some n, then $n = 0$. Similarly, for negative $a \neq -1$, if $a^n = 1$, then $n = 0$.

Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^n$ where n is a positive integer. By sequential definition of continuity, or by looking at f as the composition of multiplication function (which is continuous) with a type of diagonal function (taking x to $(x, \ldots, x)$ in $\mathbb{R}^n$, say) or by induction, one can show that f is continuous. From here it follows that polynomials and rational polynomials (wherever defined) are continous.

Next we observe the following inequalities of exponentiating to integer powers:

$$0 \leq a_1 \leq a_2, \text{ then } 0 \leq a_1^n \leq a_2^n \text{ for } n \geq 0,$$

$$0 \leq a_1 \leq a_2, \text{ then } 0 \leq a_2^n \leq a_1^n \text{ for } n \leq 0,$$

$$a \geq 1, 0 \leq n_1 \leq n_2, \text{ then } 1 \leq a^{n_1} \leq a^{n_2},$$

$$0 \leq a \leq 1, 0 \leq n_1 \leq n_2, \text{ then } a^{n_2} \leq a^{n_1} \leq 1.$$

One important equality that helps in proving the above is the following equality for any n, x, y

$$x^n - y^n = (x - y)(x^{n-1} + yx^{n-2} + \cdots + y^{n-2}x + y^{n-1}).$$

Observe that for positive x, y, the second term is positive. So, the sign of $x^n - y^n$ is the same as that of $x - y$.

Using these, observe that, for $a > 1, f = x^n$ takes the interval $[0, a]$ to $[0, a^n]$. Since $a \leq a^n$, by the intermediate value theorem, there is an $x \in [0, a]$ such that $x^n = a$. For $0 < a \leq 1$, we see that f takes $[0, 1]$ to iteself and since $0 < a \leq 1$, there is a point $x \in [0, 1]$ such that $x^n = a$. Thus, for positive a, the nth root exists. In a similar manner, one can show similar properties for negative a and then show that nth root of any a exists for n odd.

In the above, the nth root of a is that number x such that $x^n = a$. So, we can define rational powers of any number. It is clear that the nth root is unique upto a sign (this fails in the complex system). It is conventional to take the positive nth root for n even.

Given positive a, b, the nth root of ab is the product of their individual nth roots by uniqueness. So, the arithmetic properties listed above for integer powers is true for rational powers as well. I will prove just one of them and leave the rest for the reader. Given positive a and rationals $\frac{p_1}{q_1}, \frac{p_2}{q_2}$, their sum is $r = \frac{p_1q_2+p_2q_1}{q_1q_2}$, so a^r is the q_1q_2th root of $a^{p_1q_2+p_2q_1}$ which is clearly $a^{\frac{p_1}{q_1}}a^{\frac{p_2}{q_2}}$ (use the arithmetic properties of integer powers to verify this).

Theorem. *For $a > 0$, the limit of $a^{\frac{1}{n}}$ is 1.*

Proof. For $a > 1$, the sequence $a^{\frac{1}{n}}$ is decreasing and bounded below by 1. So, a limit exists. If this limit is some $c > 1$, then

$$\forall n, c \leq a^{\frac{1}{n}} \implies c^n \leq a.$$

However, the sequence c^n is unbounded as the consecutive difference is at least $c - 1 > 0$ and by the Archimedian property (which says that the natural numbers are unbounded, and therefore any positive multiple of the natural numbers is unbounded), increases without bound. Therefore, this limit must be 1. For $a < 1$, consider the sequence $a^{\frac{-1}{n}}$. Since $1/a > 1$, this sequence has limit 1 and therfore the limit of $a^{\frac{1}{n}}$ is also 1. $\square$

Observe that the four inequalities above hold for rational exponents as well. A part of the proof is given in the next subsection. Also, for positive $a \neq 1$, if $a^{\frac{p}{q}} = 1$, then raising to the qth power gives $a^p = 1$ which gives $p = 0$. A similar thing holds for negative a, but in that case we need q to be odd.

If a sequence of rational numbers $\{\frac{p_n}{q_n}\}_{n\geq 0}$ converges to 0, then for every $k \in \mathbb{N}$, there is an $N_k \in \mathbb{N}$ such that beyond N_k,

$$\left|\frac{p_n}{q_n}\right| \leq \frac{1}{k}.$$

By comparing powers of $a > 0$, since both $a^{\frac{1}{n}}$ and $a^{\frac{-1}{n}}$ converge to 1 we conclude that

$$\lim_{n\to\infty} a^{\frac{p_n}{q_n}} = 1.$$

3.2 Raising to irrational powers

It is natural to ask about raising a to irrational powers. One thing we can do is to say that for an irrational x, a^x is defined to be the limit of a^{r_n} where $\{r_n\}$ is a sequence of rationals converging to x. Using the archimedian property, it is possible to show that the rationals are dense in $\mathbb{R}$.

Lemma. *Given $x \in R$, there is a sequence of rational numbers converging to x.*

Proof. Given $x < y \in \mathbb{R}$, we show that there is a rational number between them. What we do is to find an rational number smaller that $y - x$ and then add it a number of times so that it falls between them. By the Archimedean property, there is an n such that $n(y - x) > 1$. There is also an m such that $m \leq nx < m + 1$. Combining these two, we see that $x \leq \frac{m+1}{n} < y$. Given this, we can recursively find rational numbers converging to x. $\square$

So, we can find a rational sequence converging to x. Next we need to show that the said limit exists and does not depend on the sequence of rationals we use to approximate x.

Lemma. *The definition of a^x above is well defined, i.e. the limit exists and is independent of the sequence $\{r_n\}$.*

Proof. Let $\{r_n\}, \{s_n\}$ be rational sequences converging to x. Suppose a^{s_n} converges. The sequence $r_n - s_n$ is a rational sequence converging to zero, therefore $a^{r_n - s_n}$ converges to 1. Together with the fact that a^{s_n} converges, it is clear that a^{r_n} is also convergent with the same limit as a^{s_n}. Thus, it suffices to show that for at least one sequence of rationals $\{r_n\}$, the limit a^{r_n} exists. If x is rational, this shows that this new definition of a^x and the old one agree.

When $x = 0$, then we have already shown that the limit is 1, so assume $x \neq 0$. Assume $a > 1, x > 0$ and $\{r_n\}$ is an increasing sequence of positive rational numbers converging to x. We have

$$a^{r_n} - a^{r_m} = a^{r_m}(a^{r_n - r_m} - 1).$$

Now, r_n is convergent, so bounded above by some natural number M, so $a^{r_n} \leq a^M$ for every n (because $a > 1, r_n, M > 0$). We also know that the sequence $\{r_n\}$ is Cauchy, which means that, for $\delta > 0$, we can find N such that for $m, n > N, |r_n - r_m| < \delta$.

Given $\epsilon > 0$, find an integer k such that $a^{\frac{1}{k}}$ is within ϵ of 1. Then find a suitable N for $\delta = \frac{1}{k}$. It follows that

$$0 < a^{r_n} - a^{r_m} < \frac{a^M \epsilon}{2}, n > m > N.$$

Thus, the sequence a^{r_n} is Cauchy, hence convergent. Furthermore, because $a > 1, r_n > 0$ each $a^{r_n} > 1$, therefore the limit is also greater than 1.

For $x < 0$, we take a decreasing negative sequence $\{r_n\}$. Then $a^{r_n} = \frac{1}{a^{-r_n}}$. Since the denominator converges to a non zero number (because $-r_n$ converges to $-x > 0$), this sequence converges to a non zero number. When $a < 1$ and x is arbitrary with $\{r_n\}$ converging to x, look at a^{-r_n}. Since $1/a > 1$, this sequence converges to a non zero number, therefore a^{r_n} also converges to a non zero number. Therefore, a^x is well defined. □

For positive x, we define $0^x = 0$.

Having defined a^x as a limit, the arithmetic properties listed above hold for arbitrary powers. For if x, y are irrationals, with $\{r_n\}, \{s_n\}$ rational sequences converging to x, y respectively, then

$$a^{r_n + s_n} = a^{r_n} a^{s_n}$$

and taking limits gives us the desired equality. In a similar manner, the other equalities are true.

Suppose for some positive $a \neq 1, x \in \mathbb{R}$, we have $a^x = 1$. Take a rational sequence $\{r_n\}$ converging to x. If $x \neq 0$, there is a rational $r_0 > 0$ and an N such that for $n \geq N$, $|r_n| > r_0$. For such n, assuming $x > 0$, we have $r_n > r_0$ and

- if $a > 1$, then $a^{r_n} > a^{r_0} > 1$ (because $a \neq 1$)
- if $0 < a < 1$, then $a^{r_n} < a^{r_0} < 1$.

So, the limit can't be 1. If $x < 0$, then observe that $a^{-x} = (\frac{1}{a})^x = 1$, so again the limit can't be one. Therefore we must have $x = 0$. In conclusion, for a positive $a \neq 1$, $a^x = 1$ if and only if $x = 0$. Note that a similar statement holds for negative a, but care should be taken while taking rational powers. Moreover, it is not clear as to whether arbitrary powers of negative numbers exist.

3.3 The limit $\lim_{x \to 0} a^x$ for $a > 0$

Having defined exponents to irrational powers, it is natural to ask whether the limit in the title is 1. After all, it is true for rational sequences and the exponents are defined by rational approxiamtions. First we prove the following

Theorem. *For $a > 0, \epsilon > 0$, there exists $\delta > 0$ such that for $r \in \mathbb{Q}$, with $|r| < \delta$, $|a^r - 1| < \epsilon$.*

Proof. For $\epsilon > 0$, choose $N \in \mathbb{N}$ such that $|a^{\frac{1}{n}} - 1| < \epsilon$ and $|a^{\frac{-1}{n}} - 1| < \epsilon$ for $n \geq N$. Take $\delta = \frac{1}{N}$, then if $r \in \mathbb{Q}$ such that $|r| < \delta$, we see that $|a^r - 1| < \epsilon$ as required. □

Corollary. *Let $r \in \mathbb{Q}$ and $\{r_n\}$ be a rational sequence converging to r, then $\lim_{n \to \infty} a^{r_n} = a^r$.*

Proof. For $\epsilon > 0$, there is a $\delta > 0$, such that for $s \in \mathbb{Q}$ with $|s| < \delta, |a^s - 1| < \epsilon$. For such a δ, there is an N such that for $n \geq N$, $|r - r_n| < \delta$. So,

$$|a^r - a^{r_n}| = a^r |1 - a^{r_n - r}| < a^r \epsilon.$$

The corollary is now clear. □

Theorem. *The inequalities mentioned above hold for arbitrary powers.*

Proof. I will do just one of them, the others are dealt with similarly. Let $a > 1$. If $n_1 \geq n_2$ are two integers, then $n_1 - n_2$ is a non negative integer, so $a^{n_1 - n_2} \geq 1$ which means that $a^{n_1} \geq a^{n_2}$. Now, if $r = \frac{p}{q}$ is a positive rational number, then $a^r \geq 1$ for it is the qth root of $a^p \geq 1$. So, if $r_1 \leq r_2$ are rational numbers, then, as with the case of integers, $a^{r_1} < a^{r_2}$.

Finally, let x be a positive real number. Choose a sequence of positive rationals, say $\{r_n\}$, converging to x. Then we have that for each n, $1 \leq a^{r_n}$. Taking limits, we see that $1 \leq a^x$. Given real numbers $x < y$, then it follows from what has been done above that $a^x < a^y$.

All other cases are done similarly. For example, for $a < b$, it is clear that if $r > 0$ is a rational number, then $a^r < b^r$. Now, given $x > 0$, we can take a sequence of positive rationals converging to x and then compare limits. Alternatively, we could just look at $\frac{b}{a} > 1$ and then use the arithmetic properties of exponentiation (which was proven to hold for arbitrary powers). □

Theorem. *For $a > 0, \lim_{x \to 0} a^x = 1$.*

Proof. Let $\{x_n\}$ be a sequence of real numbers. For $\epsilon > 0$, there is a $\delta > 0$ such that for rational r, if $|r| < \delta$, then $|a^r - 1| < \epsilon$. Let $r_0 > 0$ be a rational smaller than δ. There is an N such that for $n \geq N$, $|x_n| < r_0$. Assume $a > 1$, then for $n > N$, we have $-r_0 < x_n < r_0$.

If $x_n > 0$, then

$$1 < a^{x_n} < a^{r_0},$$

and if $x_n < 0$, then

$$a^{-r_0} < a^{x_n} < 1.$$

If $x_n = 0$, then clearly $a^{x_n} = 1$. So, it is clear that for $n \geq N, |a^{x_n} - 1| < \epsilon$. Thus,

$$\lim_{x \to 0} a^x = 1.$$

If $a < 1$, then some of these inqualities flip directions, but the end result is the same. □

3.4 Continuity

We have defined a^x for any real x, positive a. There are two variables here a and x. So, we have two functions, and it is reasonable to expect them to be continuous.

First we look at a^x with $a > 0$ fixed and x varying over $\mathbb{R}$. Fix $x \in \mathbb{R}$, for $y \in \mathbb{R}$, we have

$$a^x - a^y = a^x(1 - a^{y-x})$$

by the arithmetic properties of exponentiation. Since the $\lim_{x \to 0} a^x = 1$, we can make y very close to x so that the second factor is small. Further, since x is fixed, the constant factor a^x doesn't matter and it is clear that $a^x - a^y$ can be made very small. Therefore a^x is a continuous function on $\mathbb{R}$.

Next we look at x^r for fixed r and x varying over the positive real line. We have seen that x^n is a continuous function for integer n. We can also show that x^n is continuous by using the following factorization

$$x^n - y^n = (x - y)(x^{n-1} + yx^{n-2} + \cdots + y^{n-2}x + y^{n-1}).$$

Since continuity is a local condition, that is only points nearby matter, we can bound the second term, by say $y = x + 1$. That is for y within 1 of x, each $y^j < (x + 1)^j$ (when $x, y > 0$), so the whole sum is bounded by some number. The first factor can be made arbitrarily small.

However, such a factorization doesn't exist (at least, we won't have finite sums), for arbitrary r. The reason the above factorization works is that we start with $(x - y)$ and multiply by suitable terms (the first term in the second factor should be x^{n-1} this subtracts yx^{n-1}, so we add yx^{n-2} to the second term and so on). Conveniently, as the powers of x decrease there comes a stage when the next term is added, we get y^n and we don't have to look further.

However, for $r \neq 0$, we can try is to remove x^r outside. So, we have

$$x^r - y^r = x^r(1 - \frac{y^r}{x^r})$$

(which is possible because of the arithmetic properties proved earlier). Here we have two cases, $x = 0$ and $x \neq 0$. If $x \neq 0$, then as y goes closer to x, $\frac{y}{x} \to 1$. So, we have two things we need to show for continuity:

(i) $\lim_{x \to 1} x^r = 1$

(ii) $\lim_{x \to 0} x^r = 0, r > 0$.

So, in order to prove that x^r is continuous on $\mathbb{R}_{\geq 0}$ we just have to show that it is continuous at 0 and 1 (continuity at 0 is needed for $r > 0$ only).

We will prove (ii) now. Let $r > 0$ and $\{x_n\}$ be a sequence of positive numbers converging to 0. Given $k \in \mathbb{N}$, there is an N such that for $n \geq N, x_n < \frac{1}{k} \implies x_n^r < \frac{1}{k}^r$. Now it is clear that if a sequence $\{a_n\}$ goes to infinity, then $\{\frac{1}{a_n}\}$ goes to 0. So, it is sufficient to show that the sequence $\{k^r\}$ goes to infinity. This requires a proof. We know that it is an increasing sequence. Suppose it is bounded, then there is an M such that $k^r \leq M$ for every $k \in \mathbb{N}$. Raising to the power $\frac{1}{r}$ (since $r \neq 0$, this is possible), the inequality remains the same (everything is positive) and we have a bound on the integers which is a contradiction. Therefore the sequence $\{k^r\}$ is unbounded and increasing. It follows that (ii) is true. Later, we will have a more general way to prove both (i) and (ii).

4 Differentiation

Given a function f on some open set to $\mathbb{R}$, we say that f is differentiable at x if

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

exists. The reason for f to be defined on open sets is so that we can talk about points near x. It is still possible to define the derivative at end points of a closed interval, say, by talking about left handed or right handed derivative and taking h to be exclusively positive or negative. Once again, sums and products of differentable functions are differentiable and so is the quotient whenever defined. Proofs are quite standard and use the properties of limits.

4.1 The binomial expansion and polynomials

Here I show that polynomials are differentiable. Since sums and scalar multiples of differentiable functions are differentiable, it is sufficient to show that x^n is differentiable.

Theorem. *For a positive integer n, we have $(a+b)^n = \sum_{i=0}^{i=n} \binom{n}{i} a^i b^{n-i}$ where a, b are any real numbers and $\binom{n}{i}$ is the usual binomial coefficient.*

Proof. Proof is quite standard and employs induction. □

Corollary. *$\frac{dx^n}{dx} = nx^{n-1}$ for any $n \geq 0$ and all $x \in \mathbb{R}$. As a consequence, we have that the polynomials are infinitely differentiable (although eventually all derivatives are zero).*

Proof. For $n = 0$, it is obvious. For $n \geq 1$, we have

$$(x+h)^n = x^n + nhx^{n-1} + h^2 p(x)$$

where $p(x)$ is the rest of the expansion given by the binomial theorem. From here it is clear that the derivative of x^n is indeed nx^{n-1} on all of $\mathbb{R}$. □

Note that I should have been more precise by saying that x^n is defined on all of $\mathbb{R}$ and the expansion is true for any fixed x. Since the derivative at some fixed but arbitrary x is as above, it is the same everywhere. We want to show a similar result for any r. What was essential was that we can write $(x+h)^n = x^n + nhx^{n-1} + h^2 p(x)$, that is $(x+h)^n - x^n = nx^{n-1}h + R$ where R is the remainder term and satisfies

$$\lim_{h\to 0} \frac{R}{h} = 0.$$

Notice that the left hand side is a function of h, not x as x is fixed. The right hand side has two parts, the first of which is a linear function of h. So, the right hand side is a linear approximation of the left side and R is the error term which, intuitively, goes to zero faster than h. This is the idea in defining derivatives in higher dimensions, which is to approximate the difference by a linear map (which will be a matrix) and an error term that goes to zero faster than h (which will be a vector). What we want is to do the same for $(x+h)^r$ for any r.

4.2 Approximation for rational powers

We will show that given a rational r and $x > 0$,

$$(x+h)^r = x^r + rx^{r-1}h + R$$

where R is such that $\lim_{h\to 0} \frac{R}{h} = 0$. First we show this for $r = \frac{1}{n}$. We may take $x = 1$ as if it is true for this case, it is true in general as we can take x^r outside and replace h by h/x. The nth root of $(1+h)$ should be close to 1. So, we will take it to be $(1+t)$.

Then

$$(1+t)^n = 1 + nt + t^2 P$$

where t^2P is the rest of the terms. This is equal to $1+h$. Ideally, since h is small, t should also be small, so the contribution of t^2 and higher terms is negligible, so it is natural to expect t to be close to $\frac{h}{n}$. So, we take $t = \frac{h}{n} + R$, where R is the remainder.

$$(1 + \frac{h}{n} + R)^n = (1+R)^n + n\frac{h}{n}(1+R)^{n-1} + \binom{n}{2}\frac{h^2}{n^2}(1+R)^{n-2} + \cdots = 1+h.$$

Cancelling terms, we see that

$$Rp_0(R) + hRp_1(R) + \binom{n}{2}\frac{h^2}{n^2}(1+R)^{n-2} + \cdots = 0$$

where $p_0(R), p_1(R)$ are polynomials in R with non zero constant term. Note that R is a function of h, so varies with h but the equality is true for every h. Also, the terms hidden in "$\ldots$" are multiples of powers of h^3 and higher.

This is true for every h, so one can divide by any power of h and the limit as h tends to 0 should exist and should be zero. Think of it this way, given any sequence of h going to zero, the ratio is zero always, so the ratio "can be made" as close to zero as possible and hence the limit should exist and be zero. Obviously, as h tends to zero, t should tend to zero, and hence R tends to zero. This is the case when we divide by the 0th power of h. Next, divide the last equation by h. Then, the first term is

$$\frac{R}{h}p_0(R)$$

and the second term is

$$Rp_1(R)$$

and all other terms are polynomials in h and R. In the limit, $h \to 0$, $R \to 0$, so the higher terms all tend to zero. So, we have

$$\lim_{h\to 0}\frac{\cdots}{h} = 0, \lim_{h\to 0} Rp_1(R) = 0, \lim_{h\to 0} p_0(R) = p_0(0) \neq 0.$$

The third term too goes to zero, I just haven't typed it. So,

$$\lim_{h\to 0}\frac{R}{h}p_0(R) = 0 \implies \lim_{h\to 0}\frac{R}{h}\frac{p_0(R)}{p_0(R)} = 0.$$

The last implication is true because $\lim_{h\to 0} p_0(R) \neq 0$. Thus, we have what we wanted

$$\lim_{h\to 0}\frac{R}{h} = 0.$$

You may ask why we don't continue further and divide by h^2. We can do this, but in the end it is wasted effort. The total limit after dividing by h^2 is indeed 0. As before, we can ignore the terms in "$\ldots$". This time, for the third term we have

$$\lim_{h\to 0}\binom{n}{2}\frac{h^2}{n^2h^2}(1+R)^{n-2} = \binom{n}{2}\frac{1}{n^2}.$$

The problem is with the first and second terms. Here too, we can ignore higer powers of R in p_0, p_1. Specifically, we just have to care about the constants from these two polynomials.

The second term actually goes to 0 for it is

$$\lim_{h\to 0}\frac{hRp_1(R)}{h^2} = \lim_{h\to 0}\frac{R}{h}p_1(R) = 0$$

for the limit of $p_1(R)$ as $h \to 0$ exists and is the constant term of p_1.

Now, the first term. We obtained $Rp_0(R)$ as $(1+R)^n - 1$, so $Rp_0(R) = nR + \binom{n}{2}R^2 + \ldots$. Now, upon computing limits, we get

$$\lim_{h\to 0}\frac{R}{h^2} = -\binom{n}{2}\frac{1}{n^3}.$$

The limit R/h^2 exists because all other terms in the original expansion had valid limits. One may compute further, but it is pointless. We have what we want, a linear approximation with an error term that goes to zero fast enough. Computing higher rates doesn't give us anything beyond this linear approximation.

So, the first part of our work is done. We have shown what we wanted for $r = \frac{1}{n}$. Next we show the same for $n = -1$. Let $(1+h)^{-1} = 1 - h + R$, then

$$1 - h + R + h - h^2 + Rh = 1 \implies \frac{R}{h} = \frac{h}{1+h}.$$

Clearly, the limit of $\frac{R}{h}$ is then 0. Now, for any negative integer $-n$ with $n > 0$, we have $(1+h)^{-n} = (1-h+R)^n$ where $\lim_{h\to 0} \frac{R}{h} = 0$. Using the result for integers, we get

$$(1+h)^{-n} = (1 - nh + nR + R')$$

where R' is the rest of the polynomial which have terms of the form $(-h+R)^j$ for $j \geq 2$. It is clear that

$$\lim_{h\to 0} \frac{(-h+R)^j}{h} = \lim_{h\to 0} (-1 + \frac{R}{h})(R-h)^{j-1} = 0$$

for $j \geq 2$. Further, $\lim_{h\to 0} \frac{-nR}{h} = 0$. So, even for negative integers,

$$(1+h)^{-n} = 1 - nh + R$$

where R is such that $\lim_{h\to 0} \frac{R}{h} = 0$.

Now for a general rational number. Let $r = \frac{p}{q}$ where q is positive. Then, we have the first approximation

$$(1+h)^{\frac{1}{q}} = 1 + \frac{1}{q}h + R.$$

Next we raise it to the pth power and using the result for integers (both positive and negative), we have

$$(1+h)^{\frac{p}{q}} = 1 + \frac{p}{q}h + pR + R'$$

where R' is such that

$$\lim_{\frac{1}{q}h+R\to 0} \frac{R'}{\frac{1}{q}h + R} = 0.$$

However, in the limit $h \to 0$, we also have $\frac{1}{q}h + R \to 0$, so we have

$$\lim_{h\to 0} \frac{R'}{\frac{1}{q}h + R} = \lim_{h\to 0} \frac{R'}{h} \frac{1}{\frac{1}{q} + R} = 0.$$

Since the second factor has limit q, we see that $\lim_{h\to 0} \frac{R'}{h} = 0$. Now, it is clear that for any rational $\frac{p}{q}$, we have

$$(1+h)^{\frac{p}{q}} = 1 + \frac{p}{q}h + R$$

with $\lim_{h\to 0} \frac{R}{h} = 0$ (this R is the $pR + R'$ from above).

Thus, x^r is differentiable with derivative rx^{r-1} for r rational and x on the positive real line. Immediately we see that these functions are infinitely differentiable on the positive real line as well. Thus, x^r is continuous on $[0,\infty)$ (on $(0,\infty)$ when $r < 0$) and differentiable on $(0,\infty)$. To calculate the derivative at 0, we need to find

$$\lim_{h\to 0^+} \frac{h^r}{h} = \lim_{h\to 0^+} h^{r-1}.$$

This limit exists for $r \geq 1$, and is 0 for $r > 1, 1$ for $r = 1$. For $r < 1$, this limit does not exist.

4.3 Approximation for irrational powers

Naturally, we would like to show similar results as above for irrational powers. Let r be any real numbers and $\{r_n\}$ be a rational sequence converging to r. Then, we have

$$(1+h)^{r_n} = 1 + r_n h + R_n.$$

Since $R_n = (1+h)^{r_n} - 1 - r_n h$, it is clear that R_n converges to some R and we have

$$(1+h)^r = 1 + rh + R.$$

We need to show that $\lim_{h\to 0} \frac{R}{h} = 0$. However this is not as easy as it seems. What we want to show is

$$\lim_{h\to 0} \lim_{n\to\infty} \frac{R_n}{h} = 0.$$

It is tempting to exchange the limits and call it a day. Why would we think that this would work? Well, both limits seperately exist, and the whole limit exists in one order (h first and then n), so the situation is nice and shouldn't it work? No. Even in such nice situations, we can have problems. Consider the following array

$$\begin{array}{cccccccc}
1 & 1 & 1 & 1 & \dots & \to & 1 \\
0 & -1 & -1 & -1 & \dots & \to & -1 \\
0 & 0 & 1 & 1 & \dots & \to & 1 \\
0 & 0 & 0 & -1 & \dots & \to & -1 \\
\vdots & \vdots & \vdots & \vdots & \dots & \dots & \dots \\
\downarrow & \downarrow & \downarrow & \downarrow & \dots & & \dots \\
0 & 0 & 0 & 0 & \dots & \to & 0
\end{array}$$

Both horizaontal and vertical limits exist seperately. Vertical, then horizontal limit exists and is 0, but the limit with limits in the other order doesn't exist.

They key thing is that these R_n are not arbitrary numbers, they are very good approximations to certain differences. Further, the functions x^{r_n} are very good approximations to x^r. Since we already know the derivative of x^{r_n}, it is natural to ask whether the individual derivatives converge to the derivative of x^r. However, this is not true always. The functions may converge fastly in some regions and slowly in others.

This mismatch of rates of convergence means that the derivatives need not converge, for in order to compute the derivative, we need to know the function near a point at the same time. If these functions don't converge at a uniform rate (uniform in the sense of the xs), then while approximating x^r by x^{r_n}, we may be using one n for x and another for $(x+h)$ so that both are equally close to the values $x^r, (x+h)^r$. So, if they don't converge uniformly, we might have problems. This doesn't mean that if they do converge uniformly, then the derivative converges, for the individual derivative may also behave very badly. They may not converge to a function at all (this is not the case for the functions we are considering), or they themselves may have a mismatch in rates of convergece, which would mean that we may not be able to get hold of one function to which the derivatives converge. We will come back to this later.

5 Rolle's theorem and mean value theorem

Suppose we have a continuous function f on a closed interval $[a,b]$. There is a thorem called the Bolzano-Weierstrauss theorem which, using the completeness axiom, shows that any bounded sequence has a convergent subsequence. Using this, one can show that f attains a maximum and a minimum on this interval.

Suppose f is differentiable and attains an extremum at some $c \in (a,b)$, then one sees that the difference quotients are positive from one side of c and negative from the other and since these quotients converge to $f'(c)$, it follows that $f'(c) = 0$.

Theorem. ***(Rolle's theorem)*** *With f as above, if $f(a) = f(b)$, then there is a point $c \in (a,b)$ with $f'(c) = 0$.*

Proof. Since $f(a) = f(b)$, either f is constant, in which case the theorem is obvious, or a maximum or a minimum occurs in (a,b). For if both maximum and minimum occur at the end points, then f is constant. Since the derivative at the extremum lying in (a,b) is zero, the theorem follows. □

We have the following generalization of Rolle's theorem.

Theorem. *(**Mean value theorem**) With f as above, there is a point $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.*

Proof. This theorem is a simple application of Rolle's theorem. We want to level $f(a), f(b)$, and apply Rolle's thorem. What we do is we subtract the line joining the points $(a, f(a))$ and $(b, f(b))$. Here I have skipped some steps and you are to think of the previous statement as an intuition for it doesn't make sense to subtract a line. Further, the subtraction is happening in the graphs of f and the line. Intuitively, we will define a function that gives the vertical distance of f from said line. This is easily achieved by setting

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a} x.$$

Now g is continuous and $g(a) = g(b)$. From Rolle's theorem, the mean value theorem follows. □

6 Uniform convergence and derivatives

Previously I had mentioned sequences of functions that converge at the same rate to another function everywhere. We make this precise here.

Definition. *Given a function f on a set $A \subset \mathbb{R}$, we define its sup-norm by $\|f\| = \sup_{x\in A} |f(x)|$ provided it exists.*

Note that the sup norm exists for continuous functions on compact sets as continuous functions on compact sets attain a maximum. Of course, we can define such a norm on any set. This is a measure of the "size" of f. Observe that if we multiply f by a real number α, then $\|f\|$ is multiplied by $|\alpha|$. Further, since the triangle inequality holds in $\mathbb{R}$, it holds in the space of functions (one can add and scale functions, which means that the set of functions forms a real vector space), i.e., given f, g, we have

$$\|f + g\| \leq \|f\| + \|g\| .$$

Definition. *Given a sequence of functions $\{f_n\}$ on a set $A \subset \mathbb{R}$, we say that f_n converge uniformly to a function f if given $\epsilon > 0, \exists N \in \mathbb{N}$ such that for $n \geq N, \|f - f_n\| < \epsilon$.*

This definition captures what we mean by converging uniformly. Given $\epsilon > 0$, the N we get so that f_n is close to f is independent of x, it is close to f everywhere at once. Naturally, if we have differentiable functions on closed intervals converging uniformly, we would expect the derivatives to converge too. But this is not true. The derivatives require much stronger conditions to converge. However, it is true that if f_n converge uniformly to f and are all continuous, then f is continuous.

Theorem. *Let f_n be a sequence of continuous functions on $[a, b]$ and let them converge uniformly to some function f, then f is continuous.*

Proof. Take $\epsilon > 0$. There is an N such that for $n \geq N, \|f - f_n\| < \epsilon/3$. Take one such N. For f_n, there is a $\delta > 0$ such that if $|x - y| < \delta$, then $|f_N(x) - f_N(y)| < \epsilon/3$. Now, for the same δ, if $|x - y| < \delta$

$$|f(x) - f(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| < \epsilon.$$

Therefore f is continuous. □

Derivatives are not as well behaved, for example consider the sequence $f_n = n^{-1/2} \sin(nx)$, defined on $[0, 1]$ say. Then f_n converge uniformly to $f \equiv 0$. Both f_n and f are differentiable, with $f' \equiv 0$, $f_n' = n^{1/2} \cos(nx)$. Observe that f_n' don't converge anywhere. As mentioned earlier, the problem is that the derivatives don't converge or are not converging at the same rate.

Further another problem we may have is that f may not be differentiable. Consider the functions $f_n = \sqrt{\frac{1}{n^2} + x^2}$ on $[-1, 1]$. Each f_n is differentiable and since

$$|x| < f_n(x) < |x| + \frac{1}{n}$$

the f_n converge uniformly to $|x|$. However, $|x|$ is not differentiable at zero. A more dramatic example is the Weierstrauss function which is given by a series whose partial sums are differentiable, but the Weierstrauss function is nowhere differentiable (but continuous everywhere).

So, uniform convergence of functions doesn't tell whether their limit is differentiable and if so, whether the derivative of the limit is the limit of their derivatives. What about the converse? Surely requiring the derivatives to converge uniformly to some function seems like a stronger condition. We have the following theorem.

Theorem. *If f_n is a sequence of differentiable functions on $[a, b]$ such that $\lim_{n\to\infty} f_n(x_0)$ exists (and is finite) for some $x_0 \in [a, b]$ and the sequence f'_n converges uniformly on $[a, b]$, then f_n converges uniformly to a function f on $[a, b]$, and $f'(x) = \lim_{n\to\infty} f'_n(x)$ for $x \in [a, b]$.*

Proof. This proof is taken from Bartle and Sherbert's Introduction to Real analysis Chapter 8 Section 2.

So, the derivatives converge uniformly and $f_n(x_0)$ is also convergent. This means that given $\epsilon > 0$, beyong some N, $\|f'_n - f'_m\| < \epsilon$. Which means that for every x in (a, b), the difference $f'_n(x) - f'_m(x) < \epsilon$. This is the derivative of $f_n - f_m$. Now, given x, using the mean value theorem for $f_n - f_m$, we have, for some y between x, x_0

$$f_n(x) - f_m(x) = f_n(x_0) - f_m(x_0) + (x - x_0)(f'_n(y) - f'_m(y)).$$

The difference $x - x_0$ is bounded by $b - a$. Since f'_n converges uniformly, and since $f_n(x_0)$ also converges, it is clear that f_n is uniformly convergent to some f which must be continuous by the previous theorem.

Next fix a c in (a, b). We want to find the limit of $\frac{f(x)-f(c)}{x-c}$ as $x \to c$. We do this as follows. Firstly, for any x, there is a z between x, c such that

$$\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} = f'_m(z) - f'_n(z).$$

Given ϵ, find N such that for $m, n \geq N$, $\|f'_m - f'_n\| < \epsilon$. For $m, n > N$ we have

$$\left|\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}\right| < \epsilon.$$

Now, x, c are fixed and this is true for every $m, n \geq N$. So, we take the limit $m \to \infty$. This is allowed as the limit exists and m, n are independent. Therefore,

$$\left|\frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}\right| < \epsilon.$$

Let f'_n converge to a function g, then $f'_n(c)$ converges to $g(c)$, so there is a K such that for $n \geq K$, $|f'_n(c) - g(c)| < \epsilon$. Then for $M = \max\{N, K\}$, there is a $\delta > 0$ such that if $|x - c| < \delta$, then

$$\left|\frac{f_M(x) - f_M(c)}{x - c} - f'_M(c)\right| < \epsilon.$$

Combining these three inequalities, we have for $|x - c| < \delta$,

$$\left|\frac{f(x) - f(c)}{x - c} - g(c)\right| < 2\epsilon.$$

Thus, f is differentiable at c, and since it was arbitrary, it is differentiable. It is clear that $f' = g$. □

7 The derivative of x^r

So far we have shown that x^r is differentiable at all points with derivative rx^{r-1} for rational r. Now we show that this is true for any r. For a given $r \neq 0$, take a sequence of rationals $\{r_n\}$ converging to r. Restrict these functions to some compact set $[a, b]$ for some integer $0 < a < b$. Then, we know that the functions $f_n = x^{r_n}$ converge to $f = x^r$ on this interval pointwise. We show that the convergence is actually uniform. To do this, we first find the sup norm of $f_n - f_m$. This supremum is attained because the interval is compact. It may occur either at a or at b or at some point in between. We are taking any $[a,b]$ because the r_n may be negative, in which case the functions won't be defined at 0.

If the maximum occurs in (a,b), then at that point the derivative of $f_n - f_m$ is zero. Note that if this maximum is 0, then $f_n = f_m$ and the derivative is anyways 0, else the maximum is positive in which case, $f_n - f_m$ attains an extremum at that point and since it is differentiable, the derivative must be zero.

Okay, so the derivative of $f_n - f_m$ is $r_n x^{r_n - 1} - r_m x^{r_m - 1}$ and should be zero at the point of interest, say x_0. Multiplying by x_0 (don't forget, $x_0 > 0$), we have $r_n x_0^{r_n} = r_m x_0^{r_m}$.

First assume $r > 0$. Fix an $\epsilon > 0$ and assume that each $r_n > 0$ (else restrict to a subsequence), then we have

$$|x_0^{r_n} - x_0^{r_m}| = \frac{x_0^{r_m}}{r_n}|r_m - r_n|.$$

Assume that each $r_n > r - \delta > 0$ for some $\delta > 0$, then $\frac{1}{r_n} < \frac{1}{r-\delta}$.

Since each $r_n > 0$, the functions x^{r_n} are increasing and attain a maximum at b, therefore $x_0^{r_m} \leq b^{r_m}$. Now, find N_1 such that beyond $N_1, b^{r_m} < b^r + 1$. By Cauchy property of $r - n$, find N_2 such that for $m, n > N_2$, we have

$$|r_m - r_n| < \frac{\epsilon(r-\delta)}{b^r + 1}.$$

Next, we know that a^{r_n}, b^{r_n} converge to a^r, b^r respectively, so pick N_3 such that beyond N_3, $|a^{r_n} - a^{r_m}| < \epsilon$ and $|b^{r_n} - b^{r_m}| < \epsilon$. Then, it is clear that for $m, n \geq \max\{N_1, N_2, N_3\}$, we have

$$\|x^{r_n} - x^{r_m}\|_{[a,b]} < \epsilon$$

as we have taken care of the boundary points and the points in between (because the maximum is bounded by ϵ).

When $r < 0$, it is similar to the previous case except, the functions are decreasing and we have to use $a^r + 1$ as a bound. Thus, in both cases, the convergence is uniform on $[a,b]$.

Note that if the maximum occurs at $x_0 \in (a,b)$, then there is a particular condition it must satisfy, and given this condition, we bounded the difference by something depending on r_n only, and not on x_0 which depends on the pair n, m. We don't know how it varies, but we can bound the required difference by something independent of x_0 and therefore bound the overall difference. Note that even if the maximum is zero, then the derivative is zero, and the analysis still holds. So, x^{r_n} converge uniformly on $[a,b]$ to x^r.

Next we show that $r_n x^{r_n - 1}$ converges uniformly on $[a,b]$. To do this, we add and subtract the cross terms. We have

$$|r_n x^{r_n-1} - r_m x^{r_m-1}| \leq |r_n x^{r_n-1} - r_m x^{r_n-1}| + |r_m x^{r_n-1} - r_m x^{r_m-1}|.$$

The second term can be made small because r_m is bounded and x^{r_m-1} converges uniformly to x^{r-1}. The first term can be bounded because r_n is a Cauchy sequence and given any r_n, x^{r_n-1} is bounded above by either a^{r_n-1} or b^{r_n-1} and both these sequences are bounded above.

So, the derivatives converge (their pointwise convergence is obvious) uniformly to rx^{r-1} on $[a,b]$. Thus, by the theorem on uniform convergence and derivatives, it follows that $x^r$is differentiable on $[a,b]$ for any $0 < a < b$. Continuity of the same follows. Since any positive x is between some a, b as above, and differentiablility is a local property, x^r is differentable on $(0,\infty)$ with derivative rx^{r-1} for any r. Differentiablility at 0 is clear for $r > 1$. Further, it is also clear that these functions are infinitely differentiable. Phew! A big chunk of our work is over.

8 So far So good

Let us look at what we have done so far. First we defined what it means for a function to be continuous. We will leave that aside. Then we defined what it meant to raise a real number to an integral power. After that, using the intermediate value theorem, we established that nth roots exist, and using this fact, we were able to raise a positive number to any rational power. Then, it was natural to want an extension to arbitrary powers, and we did this by using sequences of rational numbers to approximate any real number. Having done that, we decided to study the analytic properties of the two functions a^x, x^r, where x is the variable and a, r are fixed constants. We showed that both of these are continuous, and found the derivative of x^r.

Along the way, we derived the binomial expansion for positive integer powers using which we were able to find a linear approximation to $(x+h)^r$. Using the linear approximation, we can easily find the derivative of x^r for rational powers. To get the derivative for irrational powers, we needed something more. The problem is that we can define rational powers directly, but irrational powers are defined in terms of limits of sequences.

To differentiate such a function, we wanted a way to interchange limits and derivatives. This, however, is not trivial. To be able to interchange limits and derivatives, a sufficient condition was the requirement of uniform convergence of the derivative. Fortunately, we do have uniform convergence. This allows us to find the derivative of x^r for any r. Of course, 0 is still a problematic point, but every other point is nice.

A couple of things should start bothering us. We have the binomial expansion for positive integers. What about other powers? After all, we do have a linear approximation that looks very much like the first two terms of a binomial expansion. Further, it is a common fact that the

$$1 + x + x^2 + \cdots = \frac{1}{1-x}, |x| < 1.$$

As you can see, this is some sort of expansion of $(1-x)^{-1}$. Is it related to the binomial expansion? Well, there are a couple of oddities. The expansion is infinite, can we do something about it? No, for if the expansion was a finite polynomial, then eventually the derivative becomes zero, whereas this is not true for $\frac{1}{1-x}$. Another issue is that the left side is convergent only on $|x| < 1$ whereas the right side is defined for $x \neq 1$. We could try some scaling of the denominator. Okay, at least we have something but why should one even want a binomial expansion?

The way I see it, binomial expansion is a computational tool. Practically, it helps in approximation and error estimation. It is similar to how one uses continued fractions as approximations of real numbers. In the case of infinite series, we can bound the terms beyond some N giving us a bound on the error in our approximation.

The other question is about the function we haven't talked about much, a^x for a fixed positive a. We have shown that it is continous. Is it differentiable? If so, what is its derivative? Answering this question leads to the rabbit hole of logarithmic functions and exponential functions.

9 Taylor series

We have a linear approximation of functions given by the derivative. Naturally, if the function is sufficiently smooth, we would expect a polynomial approximation (approximate f using f', then f' using f'' and so on). This is what Taylor expansion gives us, a polynomial approximation with a remainder term that goes to zero very fast under sufficient smoothness assumptions.

A good polynomial approximation should have matching derivatives upto some order. That is , if p is a degree n polynomial which is supposed to approximate f, then we would like to have the kth derivatives of f, p match at some point (or better, everywhere) for some range of k. We will approximate f by a polynomial in a neighbourhood of c for some c. Unless f is itself a polynomial, it would be difficult to have a better approximation, this is because if $p = a_0 + a_1(x-c) + \cdots + a_n(x-c)^n$, then the a_is are determined by the derivatives of p evaluated at c, so the derivatives can match very well at some fixed point. Note that every polynomial can be written this way by a translation of origin.

Theorem. ***(Taylor's Theorem)*** *Let $n \in \mathbb{N}$, let $I = [a, b]$, and let $f : I \to \mathbb{R}$ be such that f and its derivatives $f', f'', \ldots, f^{(n)}$ are continuous on I and that $f^{(n+1)}$ exists on (a, b). If $x_0 \in I$, then for any x in I there exists a point c between x and x_0 such that*

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}.$$

Proof. This proof and statement of the theorem are taken from the book by Bartle and Sherbert.

Let x, x_0 be given and let J denote the closed interval with endpoints x_0 and x. Define a function F on J by

$$F(t) = f(x) - f(t) - (x-t)f'(t) - \cdots - \frac{(x-t)^n}{n!} f^{(n)}(t)$$

for $t \in J$. Note that x is fixed. Then we have

$$F'(t) = -\frac{(x-t)^n}{n!} f^{(n+1)}(t).$$

If we define G on J by

$$G(t) = F(t) - \left(\frac{x-t}{x-x_0}\right)^{n+1} F(x_0)$$

for $t \in J$, then $G(x_0) = G(x) = 0$. By Rolle's theorem, we have a c between x, x_0 such that

$$0 = G'(c) = F'(c) + (n+1)\frac{(x-c)^n}{(x-x_0)^{n+1}} F(x_0).$$

Hence, we obtain

$$F(x_0) = -\frac{1}{n+1}\frac{(x-x_0)^{n+1}}{(x-c)^n} F'(c) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1},$$

which implies the stated result. $\square$

10 Finally, general binomial theorem

Now we have the equipments to derive the general binomial expansion. Historically, Newton stated this result, but did not give a proof. He probably (this is just my opinion with no knowledge on this history) extended the integer power expansion and verified it by some computations. Euler made an incomplete attempt in 1774, but the full proof had to wait for Gauss to provide it in 1812.

We can define the general binomial coefficient by

$$\binom{r}{n} = \frac{r(r-1)\dots(r-n+1)}{n!}, n \geq 0.$$

It is clear that the nth derivative of x^r is $\binom{r}{n} n! x^{r-n}$, so the Taylor polynomial is going to be the first few terms of the binomial expansion we hope to reach. Note that since x^r is infinitely differentiable, we don't have to worry about how many times we can differentiate.

We will expand $(1+x)^r$ for arbitrary r. We set

$$f(x) = \sum_{n \geq 0} \binom{r}{n} x^n.$$

First we find the radius of convergence of this series, i.e., a number R such that the series converges for x with $|x| < R$. Then we compare it with the Taylor series and show the required equality in the region of convergence.

The radius of convergence can be calculated by the root test or ratio test. In this particular case, the ratio test is far simpler. The radius of convergence R is given by

$$\frac{1}{R} = \limsup_{n\to\infty} \frac{\binom{r}{n+1}}{\binom{r}{n}} = 1.$$

So, the series converges for $|x| < 1$. We take the Taylor polynomial centred at 0. For $|x| < 1$, there is a c between 0 and x such that

$$(1+x)^r = P_n(x) + \binom{r}{n+1}(1+c)^{r-n-1}(x-0)^{n+1}$$

where $P_n(x)$ is the sum of first n terms of the series above. We set $R = \binom{r}{n+1}(1+c)^{r-n-1}(x-0)^{n+1}$.

Lemma. *For any $r > -1$, $\lim_{n\to\infty} \binom{r}{n} = 0$.*

Proof. Let k be an integer such that $-1 \leq k \leq r < k+1$, then for $n > k+1$,

$$\binom{r}{n} = \frac{r(r-1)\dots(r-k)}{n(n-1)\dots(n-k)} \frac{(r-k-1)\dots(r-n+1)}{(n-k-1)\dots(n-n+1)}.$$

Now $r(r-1)\dots(r-k)$ is a fixed constant, therefore

$$\lim_{n\to\infty} \frac{r(r-1)\dots(r-k)}{n(n-1)\dots(n-k)} = 0.$$

It suffices to show that the second term is bounded, then the required limit is zero. We look at the modulus of the second term,

$$\begin{aligned}\prod_{i=k+1}^{n-1} \left|\frac{r-i}{n-i}\right| &= \prod_{i=k+1}^{n-1} \frac{i-r}{n-i} \\ &\leq \prod_{i=k+1}^{n-1} \frac{i-k}{n-i} \\ &= \frac{1\times 2\times\dots\times(n-k-1)}{(n-k-1)\times\dots\times 2\times 1} \\ &= 1\end{aligned}$$

When $r=-1$, it is easy to see that the limit does not exist because the coefficients osscilate between ± 1. For other negative numbers, each ratio has absolute value >1, therefore the limit is not zero. $\square$

We may be able to send the binomial coefficient to zero for some r, we check if for those r, the remainder R goes to 0. Suppose $x>0$, then $0<c<x<1$ and c depends on n. But, $1<1+c<1+x<2$. Eventually, $r-n-1<0$, so for those n, we have

$$1 > (1+c)^{r-n-1} > (1+x)^{r-n-1}.$$

So, the sequence $(1+c_n)^{r-n-1}$ is bounded.

Suppose $x<0$, then $-1<x<c<0$, so $0<1+x<1+c<1$. Again for n such that $r-n-1<0$, we have $(1+x)^{r-n-1} > (1+c)^{r-n-1} > 1$. Whoops! seems like this doesn't work. Hmm. What could have gone wrong? Is the series approximation wrong? Or is it the Taylor expansion? Well, the series does converge on $(-1,1)$, so that shouldn't be wrong. Could it be the Taylor series?

The point is, once we show that the remainder term goes to zero with higher polynomials, we can take the difference between the infinite sum and the Taylor approximation, and with larger terms, this difference goes to 0, so the Taylor series would converge to $(1+x)^r$ for $|x|<1$.

11 Taylor series revisited

So, what do we have? We have a sufficiently smooth function f on some interval I. We have two points x, x_0 and want an approximation near x_0. We take the polynomial $P_n = \sum_{i=0}^{i=n} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^i$ and see that the derivatives upto nth order of P_n and f match at x_0 (in fact it is the only such polynomial) and we expect P_n to approximate f very close to x_0. What we have to worry about is the remainder $R_n(x) = f(x) - P_n(x)$.

Previously, we calculated this remainder using Rolle's theorem by considering the function $F(t) = f(x) - P_n(t)$. From here, we obtained another function G which was 0 at x, x_0 so that Rolle's theorem was applicable. The question is whether this the best estimate of R_n. We may have different expressions of the remainder term and different expressions may have different bounds. The expression we obtained before is

$$\frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}$$

and is called the **Lagrange form**. In the previous section it was observed that the remainder term written as such doesn't have an upper bound that goes to zero, or at least we weren't able to find one.

So, we have to ask whether the remainder term necessarily goes to 0 with larger polynomials. Further, are there other ways to write the remainder term so that we can find better bounds on the remainder term and apply it to the case of $(1+x)^r$? With F as above, we can apply the mean value theorem to get a point c between x, x_0 (since F is defined on the closed interval with end points x, x_0, this is possible) such that

$$F'(c) = \frac{F(x_0) - F(x)}{x_0 - x} = \frac{F(x_0)}{x_0 - x}.$$

By a simple manipulation, we see that the remainder term is now given by

$$R_n(x) = \frac{f^{(n+1)}(c)}{n!}(x-c)^n(x-x_0).$$

This is called the **Cauchy form**. Well, does this help with our approximation? No, we would still be stuck with an upper bound that goes to infinity. What else could we try? Let us start with the constant approximation, we have $R_1(x) = f(x) - f(x_0)$. The next remainder is $R_2(x) = f(x) - f(x_0) - f'(x_0)(x - x_0)$.

Now, we are lucky in that the functions we are interested are very smooth. Observe that $R_1(x) = \int_{x_0}^x f'(t)dt$ and $R_2(x) = \int_{x_0}^x f'(t) + \int_{x_0}^x (f'(t)(x-t))'dt$.

11.1 Integration

Let f be a real valued function defined on some interval $[a, b]$. The integral of f from a to b written as $\int_a^b f$ is the area under the curve f. Why do we care about this area, well it has a rich history. It is natural to ask about the area under curves, or contained in shapes, the volume under or contained in surfaces. To answer these questions, what we do is to approximate the "base" of our surface, which may be a rectangle on which a pyramid is placed, or a disc on which a cone is placed, or a line segment on which a parabola is drawn etc, by small rectangles and then compute the volume of the cuboid/rectangles with one base determined by the small rectangle and the other height determined by the height of the surface/ function we are considering at that point and add them all up and take a limiting value. Here I will consider only (Riemann) integration of real valued functions on closed intervals of $\mathbb{R}$.

Let $I = [a, b]$ be a closed interval and f a real valued function. A partition of I is a finite ordered set $P = (x_0, x_1, \dots, x_n)$ with $a = x_0 < x_1 < \dots < x_n = b$. The norm (or mesh) of P is defined by

$$\|P\| = \max\{x_1 - x_0, \dots, x_n - x_{n-1}\}.$$

If a point t_i has been selected from $I_i = [x_{i-1}, x_i], 1 \leq i \leq n$, then the points are are called tags of the subintervals I_i. A set of ordered pairs $\{([x_{i-1}, x_i], t_i)\}_{i=1}^n$ is called a tagged partition of I.

We also define the maximum M_i and minimum m_i of f over I_i. We then define the Riemann sum of f corresponding to the tagged partition above to be the sum

$$\sum_{i=1}^n f(t_i)(x_0 - x_{i-1}).$$

Similarly, we define the upper and lower sums $U(f; P), L(f; P)$ by replacing $f(t_i)$ by M_i and m_i respectively.

We say that f is integrable on I and the integral is A if the Riemann sums approach A when the partition norm decreases to 0. This is of course phrased in a similar manner to the definition of limits. It is also true that this definition is equivalent to saying that the upper sums decrease to A and lower sums increase to A.

I will state some facts and theorems without proofs. The proofs can be found in the book by Bartle and Sherbert.

1. The integral is unique when it exists.
2. If two integrable functions f, g defined on I agree at all but finitely many points, then their integrals are the same.
3. The operation of integration is linear (distributes over addition, and scales as expected when the function is scaled).
4. If a function is integrable on I, then it must be bounded.
5. Continuous functions and monotonic functions are integrable.
6. (Additivity) If f is integrable on $I = [a, b]$ and $c \in (a, b)$, then f is integrable on $[a, c]$ and $[c, b]$ with

   $$\int_a^b f = \int_a^c f + \int_c^b f.$$

   A corollary is that functions with finitely many discontinuities are integrable.

Okay, those were some facts. Now we come to one of the most important results in calculus. The fundamental theorem. Intuitively, it says that integration and differentiation are inverse operations. At first look, this result comes out of nowhere. What does the derivative have to do with integration? Derivatives are differences and ratios, whereas integration is sums. Derivative is defined at points, whereas integration is defined on whole intervals.

Observe that if we take f and consider the area under f from a to $x \in [a, b]$, then we have a function. It can be shown that this function, the area function, is continuous.

11.2 First fundamental theorem of calculus

Theorem. ***Fundamental Theorem of Calculus(First Form)*** *Suppose there is a finite set E in $[a, b]$ and functions $f, F\colon [a, b] \to \mathbb{R}$ such that:*

(a) F is continuous on $[a, b]$,

(b) $F'(x) = f(x)$ for all $x \in [a, b] \setminus E$,

(c) f is integrable on $[a, b]$.

Then we have

$$\int_a^b f = F(b) - F(a).$$

Proof. E is the set of points where the derivative of F doesn't agree with f. We can take $E = \{a, b\}$ and the general case is done by breaking the original interval and using additivity of the integral. Let $\epsilon > 0$ be given. Since f is integrable, there is a $\delta > 0$ such that if P is any tagged partition with $\|P\| < \delta$, then

$$|S(f; P) - \int_a^b f| < \epsilon$$

where $S(f; P)$ is the Riemann sum corresponding to the tagged partition P.

If the subintervals in P are $[x_{i-1}, x_i]$, then by the mean value theorem, there is a $u_i \in (x_{i-1}, x_i)$ such that

$$F(x_i) - F(x_{i-1}) = F'(u_i)(x_i - x_{i-1}), i = 1, \ldots, n.$$

Adding these terms and using the fact that $F'(u_i) = f(u_i)$, we have

$$F(b) - F(a) = \sum_{i=1}^{n} f(u_i)(x_i - x_{i-1}).$$

Since the right side is the Riemann sum, it follows that for partitions of norm less than δ,

$$|F(b) - F(a) - \int_a^b f| < \epsilon.$$

If one uses the (equivalent) definition of upper sums and lower sums converging to a point, then we see that the right side is sandwiched between the upper and lower sums and the same result follows. Since ϵ is arbitrary, the theorem is true. $\square$

Remark. If f is not defined at some point in between, we can define it to be 0 at that point and the theorem holds. However, even if F is differentiable, it is not guaranteed that its derivative f is integrable. There are functions F such that F' exists but is not integrable. See Volterra's function.

11.3 Second fundamental theorem of calculus

Definition. *If f is an integrable function on $[a, b]$, then the function defined by*

$$F(z) = \int_a^z f \text{ for } z \in [a, b],$$

is called the indefinite integral of f with basepoint a.

Theorem. *The indefinite integral F defined above is continuous on $[a, b]$. In fact if $|f(x)| \leq M$ for all $x \in [a, b]$, then $|F(z) - F(w)| \leq M|z - w|$ for all $z, w \in [a, b]$.*

Proof. Suppose $z, w \in [a, b]$ with $w \leq z$, then by additivity, we have

$$F(z) - F(w) = \int_w^z f.$$

By comparision properties of integral, we have

$$-M(z - w) \leq \int_w^z f \leq M(z - w).$$

The theorem follows. $\square$

What this theorem shows is that F is Lipschitz continuous, hence it is continous. Next we show that F is differentiable.

Theorem. ***Fundamental Theorem of Calculus (Second Form)*** *Let f be an integrable function on $[a, b]$ continous at a point $c \in [a, b]$. Then the indefinite integral F defined above is differentiable at c and $F'(c) = f(c)$.*

Proof. We will suppose $c \in [a, b)$ and consider the right hand derivative of F. Since f is continuous at c, given $\epsilon > 0$, there is a $\delta > 0$ such that if $|h| < \delta$, then $|f(c+h) - f(c)| < \epsilon$. Now, for $0 < h < \delta$, we have $F(c+h) - F(c) = \int_c^{c+h} f$. Using comparision properties of integral, we see that

$$(f(c) - \epsilon)h \leq F(c + h) - F(c) = \int_c^{c+h} f \leq (f(c) + \epsilon)h.$$

From here it follows that

$$\lim_{h \to 0} \frac{F(c + h) - F(c)}{h} = f(c).$$

Similarly, the left hand derivative of F exists for $c \in (a, b]$ and equals $f(c)$. The assertion follows. $\square$

Theorem. *If f is continous on $[a, b]$, then the indefinite integral F is differentiable on $[a, b]$ and $F' = f$ on $[a, b]$.*

11.4 Change of variables and products

When is a function integrable? Must it be continuous? No, a continuous function with finitely many discontinuities is integrable. Further, there are functions such as the Thomae's function which are discontinuous at rationals and still integrable.

Such questions led Lebesgue among others to the notion of measures. I will not go into the details here for I have already come too far off my original track. But the main result is that a function on a closed interval is integrable if and only if its discontinuities have measure zero. Intuitively, a set has measure zero if it doesn't occupy too much space, that is we can cover it by open sets in such a way that the volume (or length or area) we use in such a cover may be made as small as possible.

Given this result, we can compose two integrable functions to get another integrable function.

Theorem. ***(Composition theorem)*** *Let f be integrable on $[a, b]$ with $f([a, b]) \subseteq [c, d]$ and let ϕ be a continuous function on $[c, d]$. Then the composition $\phi \circ f$ is integrable on $[a, b]$.*

Proof. If f is continuous at $u \in [a, b]$, then so is $\phi \circ f$, so the discontinuities of $\phi \circ f$ are contained in the discontinuities of f, and hence have measure zero. Therefore $\phi \circ f$ is integrable. $\square$

We need ϕ to be continuous above, for there are examples where the theorem above doesn't hold. To find a counterexample, we need to find ϕ and f such that the $\phi \circ f$ has too many discontinuities. The indicator of rationals is a non integrable function. So, if we let f be the Thomae's function, then it has value $1/q$ for rationals p/q and 0 at irrationals. This is close to an indicator function, but is integrable. Now, we let $\phi(x) = 1$ for $x \neq 0$ and 0 otherwise.

Clearly, ϕ, f are integrable on $[0, 1]$ but $\phi \circ f$ is not integrable on $[0, 1]$ as it is the indicator function.

Theorem. ***(Product theorem)*** *If f, g are integrbale on $[a, b]$, then the product fg is also integrable on $[a, b]$.*

Proof. Let $\phi(t) = t^2$ defined everywhere. The image of f lies in some bounded set (because f is integrable, it must be bounded), so we can take the domain of ϕ large enough to contain the image of f. By the composition theorem, f^2 is integrable. Similarly, $(f + g)^2, g^2$ are integrable. Since we can write fg as

$$fg = \frac{1}{2}\Big[(f + g)^2 - f^2 - g^2\Big],$$

it follows that fg is integrable on $[a, b]$. □

11.5 Integration by parts

Using the fundamental theorm, we can integrate a derivative to get back the original function. A rather neat thing happens when we apply this to the product rule. If f, g are differentiable on $[a, b]$, then by the product rule we have

$$(fg)' = f'g + fg'.$$

Integrating this should give us fg back on the left side, but on the right side it seems to allow us to transfer the derivative from f to g.

Theorem. *Let F, G be differentiable on $[a, b]$ and let $f = F'$ and $g = G'$ be integrable, then*

$$\int_a^b fG = FG(b) - FG(a) - \int_a^b Fg.$$

Proof. Since F, G are differentiable, we have

$$(FG)' = fG + Fg.$$

Since the f, g are integrable and F, G are continuous, all integrals in question exist (use the product theorem). Using the fundamental theorem of calculus, we have

$$FG\Big|_a^b = \int_a^b (FG)' = \int_a^b fG + \int_a^b Fg$$

as required. □

11.6 Integral form of the remainder

We continue with the development of the estimate of the remainder term that was hinted to be an integral.

Theorem. *Suppose that $f', \dots, f^{(n)}, f^{(n+1)}$ exist on $[a, b]$ and that f^{n+1} is integrable, then we have*

$$f(b) = f(a) + \frac{f'(a)}{1!}(b - a) + \dots + \frac{f^(n)(a)}{n!}(b - a)^n + R_n,$$

where the remainder is given by

$$R_n = \frac{1}{n!}\int_a^b f^{(n+1)}(t)(b - t)^n.$$

Proof. This is an easy consequence of integration by parts. We have

$$\begin{aligned} R(b) &= \int_a^b f' \\ &= -\int_a^b ((b - t)f')' + \int_a^b (b - t)f'' \\ &= f'(a)(b - a) - \int_a^b \left(\frac{(b - t)^2}{2!}f''\right)' + \int_a^b \frac{(b - t)^2}{2!}f'' \end{aligned}$$

Continuing this way, we arrive at the given formula. Alternatively, we can start at $R = \frac{1}{n!}\int_a^b f^{(n+1)}(t)(b-t)^n$ and integrate by parts as follows.

$$\begin{aligned}
R_n &= \frac{1}{n!} f^{(n)}(t).(b-t)^n\Big|_a^b + \frac{1}{(n-1)!}\int_a^b f^{(n)}(t).(b-t)^{n-1} \\
&= -\frac{f^{(n)}(a)}{n!}.(b-a)^n + \frac{1}{(n-1)!}\int_a^b f^{(n)}(t).(b-t)^{n-1}
\end{aligned}$$

and so on. $\square$

12 Back to the binomial series

The integral form of the remainder term does not refer to any point in between. However, if one tries to use the remainder form as it is, to our problem of finding the bound for the remainder of $(1+x)^r$ when $x < 0$, we have a problem.

I just spent a couple of frantic minutes (I am writing this part on 28th September 2019) upon finding that the natural bounds on the remainder, even in the integral form, goes to infinity. This is because, the integral form gives the remainder for the right end of the interval when we expand about the left end.

This needs to be done with care. We will now do the calculations when we expand about the right end b of $[a, b]$.

$$\begin{aligned}
R &= f(a) - f(b) \\
&= \int_a^b -f' \\
&= \int_a^b (-(t-a)f')' + \int_a^b (t-a)f'' \\
&= -(b-a)f'(b) + \int_a^b \left(\frac{(t-a)^2}{2!}f''\right)' - \int_a^b \frac{(t-a)^2}{2!}f''' \\
&= -(b-a)f'(b) + \frac{(b-a)^2}{2!}f''(b) - \int_a^b \left(\frac{(t-a)^3}{3!}f^{(3)}\right)' + \int_a^b \frac{(t-a)^3}{3!}f^{(4)}) \\
&= -(b-a)f'(b) + \frac{(b-a)^2}{2!}f''(b) - \frac{(b-a)^3}{3!}f^{(3)}(b) + \int_a^b \frac{(t-a)^3}{3!}f^{(4)})
\end{aligned}$$

and so on. We see that the remainder is given by

$$R_n = (-1)^{n+1}\int_a^b \frac{(t-a)^n}{n!}f^{(n+1)}.$$

Now, we apply to $f = (1+x)^r$ with $a = x, b = 0$ where $-1 < x < 0$. Then, the remainder term has the following bounds

$$|R_n| \leq \int_x^0 \left|\frac{(t-x)^n}{n!}(n+1)!\binom{r}{n+1}(1+t)^{r-n-1}\right|.$$

We have

$$(n+1)\binom{r}{n+1} = (r-n)\binom{r}{n}$$

and for $-1 < x \leq t \leq 0$, we have $|t-x| < 1+t$, so

$$\frac{t-x}{1+t} < 1 \text{ for } -1 < x \leq t \leq 0.$$

Since the interval $[x, 0]$ is compact, the function $\frac{t-x}{1+t}$ attains a maximum and since $x \neq -1$, this maximum q cannot be equal to 1 and hence, must be less than 1.

So, we have

$$\left|\frac{(t-x)^n}{n!}(n+1)!\binom{r}{n+1}(1+t)^{r-n-1}\right| \leq (r-n)\binom{r}{n}q^n(1+t)^{r-1}.$$

Upon integrating, we have

$$|R_n| \leq |x|(r-n)\binom{r}{n}q^n \max\{1,(1+x)^{r-1}\}$$

where we have used the fact that since $1+x<1+t<1$, we either have $(1+t)^{r-1}<1$ or $(1+t)^{r-1}<(1+x)^{r-1}$.

Ignoring the constant the succesive ratios of $(n+1)\binom{r}{n+1}$ converge to 1. So, the radius of convergence of the series with $(n+1)\binom{r}{n+1}$ as coefficients is 1 which means that the individual terms must go to zero when evaluated at $|y|<1$. In particular, evaluating the series at q says that these bounds must go to 0 even though the coefficients themselves (e.g. the binomial coefficient) need not go to zero.

Finally, we have shown that the remainder term for the Taylor approximation around zero for $(1+x)^r$ goes to zero for $|x|<1$. Since the binomal series converges this region, and the Taylor expansion matches with the binomial series, the binomial series converges to $(1+x)^r$ for $|x|<1$. Phew! What a ride!

13 Derivative of a^x

Okay, we have dealt successfully with x^r. Now we come to a^x where a is a fixed positive number and x varies over $\mathbb{R}$. We have seen that this function is continuous. We ask whether it is differentiable. We need to compute the limit

$$\lim_{h\to 0}\frac{a^{x+h}-a^x}{h} = a^x\lim_{h\to 0}\frac{a^h-1}{h}.$$

So, we just have to find this second limit. If a^x is differentiable at 0, then it is differentiable everywhere. It is also clear that if it is differentiable once, it is infinitely differentiable.

If $a<1$, we have

$$\frac{a^h-1}{h} = \frac{a^h(1-a^{-h})}{h},$$

so the derivative exists for $a<1$ if and only if it exists for $a^{-1}>1$ because $a^h\to 1$ as $h\to 0$. We just have to show the existence of this limit for $a>1$. To prove that this limit exists for $a>1$, we observe that if we show that $\frac{a^x-1}{x}$ is increasing in x, then since it is bounded below by 0, a limit exists.

What follows is my proof, later I give an algebraic proof taken from a page mentioned in the references. Fix $r>s$ nonzero real numbers and $a>1$. Now, we want to show $\frac{a^r-1}{r}>\frac{a^s-1}{s}$. Assume $r>s>0$. Then consider the function

$$f(x) = sx^r - s - rx^s + r, x\geq 0.$$

Its derivative is (note that we have already shown x^r is differentiable for any r, so we can do this) given by

$$sr(x^{r-1}-x^{s-1}).$$

Since $r>s$, it follows that the derivative is non negative for $x\geq 1$ and is zero only at $x=1$. So, the function is non decreasing on $x\geq 1$. Since $f(1)=0$, we must have $f(a)>0$ for $a>1$. In fact, since f' has only one root, we have that f is strictly increasing to the right of 1. To the left of 1 however, the derivative is negative and since at 0 it is positive, the function (strictly) decreases from $r-s$ at 0 to 0 at 1.

A similar analysis shows the same result for any r,s. One has to be careful to flip inequalities when dealing with negatives. So, $\frac{a^x-1}{x}$ is increases as x increases.

We can also give an algebraic proof of the same as follows. First we show the result for positive integers. If n is a positive integer and $a>1$, then we are looking at a^n-1. We have the following result

$$\begin{aligned}
a^n-1 &= (a-1)(1+a+\cdots+a^{n-1}) < na^n(a-1)\\
a^n-1 &< na^{n+1}-na^n\\
(n+1)a^n-1 &< na^n\\
\frac{a^n-1}{n} &< \frac{a^{n+1}-1}{n+1}
\end{aligned}$$

So, the result is true for positive integers. Next if $r = a/b, s = c/d$ are postive rational numbers with a, b, c, d positive, then $r > s$ is the same as saying $ad > bc$. Taking $\gamma = a^{\frac{1}{bd}}$, and using the result above for positive integers, we have

$$\frac{\gamma^{ad} - 1}{ad} > \frac{\gamma^{bc} - 1}{bc}$$

or

$$\frac{a^r - 1}{s} > \frac{a^s - 1}{s}.$$

Now if r, s are two positive irrationals, then by taking sequences, we have the (weak) inequality

$$\frac{a^r - 1}{r} \leq \frac{a^s - 1}{s}.$$

Thus, the limit

$$\lim_{h \to 0^+} \frac{a^h - 1}{h}$$

exists. For $h \to 0^-$, we can replace h with some $-k$ where $k \to 0$ and see that this limit exists and is the same as the right handed limit. Since both sided limits exist and are equal, the limit exists.

We have arrived at the definition of the natural logarithm. Congratulations.

Definition. *For $a > 0$, we define the logartihm $L(a)$ to be the limit*

$$\lim_{h \to 0} \frac{a^h - 1}{h}.$$

Thus a^x is differentiable everywere with derivative $a^x L(a)$. It is common to write $L(a) = \ln(a)$, but since I haven't yet defined e, I shall avoid this. An easy exercise shows that

$$L(ab) = L(a) + L(b), a, b > 0.$$

Further, it is clear that L is non decreasing because if $a > b$, then

$$a^h > b^h, h > 0; a^h < b^h, h < 0.$$

Thus we have an non decreasing function L on the positive real line.

14 Some history

Okay, so we have defined a the logarithm. Well, not quite. We have arrived at a function L that is defined on the positive reals and satisfies $L(ab) = L(a) + L(b)$. The natural thing to do next is to explore the analytic properties of L, is it continuous?, differentiable? etc. Before doing this, I will put together some history here.

As far as my reading goes, calculus had its beginings around the time of the ancient Greece (just like almost all other sciences). Their main problems were geometrical and were centered around planar geometry specifically. They wanted to find tangents to curves, area bounded by curves. Specifically, they spoke of quadrature. To them, finding the area of a bounded figure was to find the side of a square whose area, obtained by squaring (hence quadrature) was the same as the required area. The Greeks also wanted their constructions to be done using a compass and straight-edge only, so one cannot measure exact lengths, but can measure proportions. This was a general scheme in their geometry. Using these simple tools, one can bisect a line, an angle, draw perpendiculars, take square roots, and squares, multiply and add numbers. Today, we say that the numbers so obtained form the field of constructible numbers.

However, there were many things that the Greeks couldn't do and they mainly had to do with doubling the cube (find a cube of double volume; boils down to constructing the cube root of 2), trisection of angles (some angles were fine, but others were problematic) and most famously, squaring the circle (find a square of the same area as a circle, that is to find the quadrature of the circle). They also pondered over constructing arbitrary regular polygons, but couldn't go beyond the hexagon. Later Gauss would show that the regular 17-gon is constructible.

Archimedes (288BC-212BC) and others discovered the method of exhaustion where they would approximate the figure by polygons from outside and from inside and, in retrospect, take the limit. This doesn't strictly follow the compass and straight-edge constructions, but was acceptable. Archimedes found the quadrature of the parabola and found good estimates of π among other things he used this method for.

An alternative approach was Cavalieri's (1598-1647) principle or the method of indivisibles. Simply said, it observes that curves are made by infinite points (indivisibles) and regions by lines. If two regions confined between two parallel lines are such that lines (parallel to the ends) intersect both regions in equal lengths, then the two regions have equal area. A similar version was stated for solids. Both these methods were important precursors for calculus as developed by Newton (1643-1727) and Leibniz (1646-1717). Particularly, Cavalieri's principle was later developed into a theory of infinitesimals by Torricelli (1608-1647) and Wallis (1616-1703).

Around the same time, astronomy was gaining momentum. Kepler (1571-1630), from the data gathered by Tycho Brahe (1546-1601), discovered Kepler's laws of planetary motion. There is this one law that talks about areas and this other one that says that orbits are ellipses, so it became important to study the quadrature of the ellipse. Note that conic sections have been around since Archimedes. In the area of astronomy, numbers were being observed whose magnitudes had, perhaps, never been comprehended as real by the human mind at any time before in our history. It was important to find a way to work with these numbers. It was around this time that Napier (1550-1617) came up with the natural logarithm. He wasn't the first one, logarithms were being used before him; but his logartihms were more user-friendly.

Before logarithm, there was the method called Prosthapharesis. This method used known trignometric identites to multiply numbers. Basically, given two numbers, scale them down to be between -1 and 1. Then, by looking at trignometric tables, find the sine or cosine inverses. Then, add or subtract the two numbers, take their sine or cosine and use the trignometric identities for sum and difference of trignometric functions. This method works well, but requires a lot of adding and subtracting. Note that trignometry is ancient and using large circles, it is possible to construct trignometric tables. Trignometry was motivated by problems in astrnomy regarding how to measure lengths and angles, and from problems in navigation.

Prosthaphaeresis is hard and logarithms are easier to handle. Napier wanted to simplify calculations and arrived at a version of logarithms. Napier and Briggs (1561-1630) came together to improve Napier's calculations and in the end Briggs published logarithmic tables. This became quite popular.

Next comes Pierre de Fermat (1607-1665). This guy was a giant. I would say that he was one of those giants on whose shoulders Newton stood. So, Fermat developed on previous ideas to come up with a way to find tangents and quadratures. He developed the technique of adequality to find minima and maxima of functions. This method is very similar to what we do today, although it was criticized by the likes of Descartes (1596-1650). He also found a way to evaluate the quadrature of power functions. Now, Cavalieri had discovered the method of indivisibles and successfully computed the quadrature of a line, and of parabolas. He approximated the area by rectangles in a similar way to how integrals are computed today. Wallis then did the same for other power functions. Fermat then does something better, he reduces all the previous calculations to geometric sums, and then does a limiting step which was criticized by some. By this time the quadrature of all but the hyperbola ($xy = 1$) was computed.

Then Gregorie de Saint-Vincent (1584-1667) took a major step in finding the quadrature of the hyperbola. He showed that if the ends are increased geometrically, then the areas increase arithmetically. In modern notation, he showed

$$\int_1^{ab} \frac{1}{x} = \int_1^{a} \frac{1}{x} + \int_1^{b} \frac{1}{x}.$$

A student and co-worker A. A. de Sarasa (1618-1667) noticed that this property is similar to that of the logarithm, both reduce a multiplication to an addition.

Having developed all this, it was Newton (1643-1727) who came next. Well, Newton and Leibniz. Calculus was developed, the integral and derivative were defined. Note that the fundamental theorem is not very obvious, it relates a certain area to a certain tangent. If one looks closely however, one sees that it does not come out of the blue. If one is to look at the rate of change of area, then one considers an infitesimal rectangle under a curve, compute its area and divide by its width. In the end this gives us its height, which in the limit should be the value of f at that point. Conversely, if one integrates the derivative, then one is multiplying by the width and adding all the terms, intuitively, this sum telescopes (by the way the derivative is defined) and we get the difference of the value of f at the ends.

So, anyway Calculus is born. Then comes another giant, Euler (1707-1783). Among other things he did, Euler treated the logarithm as the inverse of exponentiation (which takes a sum and gives a product). He then defined logarithm (to the base e) to be the one we have arrived at.

The history of this subject is rich. I have left out a lot of people like Roberval (1602-1675), Nicholas Mercator (1620-1687) who were influential in the development of calculus and logarithms. Moreover, history is messy, nothing was done in the first try, errors were made and were corrected by different people. We have the benefit of calculus having been developed to what it is today and are able to apply it rigorously, but very often it is important to look at the foundations of the building we are standing on, to look at who built it, how and, more importantly, why.

15 Onward

Our definition of the logartihm of a positive number a is

$$L(a) = \lim_{h \to 0} \frac{a^h - 1}{h}$$

which we arrived at by considering the derivative of a^x at 0. By the fundamental theorem of calculus, this is

$$\lim_{h \to 0} \int_1^a x^{h-1}.$$

We have already shown that the derivative of x^h is x^{h-1}, and that this function is nice enough to apply the fundamental theorem of calculus.

We prove that x^{h-1} converges uniformly to x^{-1} on $[1, a]$ as follows. Observe that if h is a decreasing sequence of positive numbers, then the function x^{h-1} monotonically decreases to x^{-1} for $x > 1$. On the other hand if h increases to 0 from the negative side, then the functions monotonically increase to x^{-1}. Now, $[1, a]$ is compact and so for monotonic sequences, the pointwise convergence implies uniform convergence by **Dini's Theorem**. Now we take the specific sequences $h_n = \frac{1}{n}$ and $h_n = \frac{-1}{n}$. For a general sequence of h converging to 0, we can use the two sequences above (which we know converge uniformly and monotonically to $\frac{1}{x}$) and bound all other functions. So, the convergence is uniform on $[1, a]$.

Next, we use the fact that under uniform convergence, limit and integral can be interchanged. Then, we have

$$L(a) = \int_1^a \lim_{h \to 0} x^{h-1} = \int_1^a x^{-1}.$$

Note that $\lim_{h \to 0} x^{h-1} = x^{-1}$ is true because we have already proved that $a^x \to 1$ as $x \to 0$. Thus, with our modern tools, we have connected the definition of logarithm to the area under the hyperbola. Furthermore, this shows that L is differentiable and its derivative is $\frac{1}{x}$. Note that we also have a proof of St. Vincent's area theorem.

16 Further properties of log and the number e

Now Euler thought of the logarithm as the inverse of exponentiation. So, he wanted a number a such that $a^{L(x)} = x$. In our modern notion, we see that the derivative of L is never zero, so it is injective. Moreover, it is looks surjective because of the way the curve, hence the area under, $\frac{1}{x}$ looks. More rigorously, by the inverse function theorem, at least local inverse exists to L. Intuitively, if $L(a)$ is given by

$$\lim_{h \to 0} \frac{a^h - 1}{h},$$

then its inverse should be given by

$$\lim_{h \to 0} (1 + hb)^{1/h}$$

(just invert and them take limit). So, Euler considers this limit, written differently,

$$\lim_{r \to \infty} \left(1 + \frac{a}{r}\right)^r.$$

Before we show that this limit exists, we explore properties of logarithm as defined above. We have already seen $L(ab) = L(a)L(b)$ for $a, b > 0$. With a little work, one can show $L(a^b) = bL(a)$ for $a > 0, b \in \mathbb{R}$. Next, we have the formula

$$L(a) = \int_1^a \frac{1}{t}.$$

So, L is an increasing function, and as $a \to \infty$, it is clear that $L(a) \to \infty$ because of the divergence of the harmonic series (basically we can get a lower bound for L using the harmonic sum). So, towards the right L is unbounded and goes of to ∞. By continuity, intermediate value theorem holds, so L takes every positive value. On the other side, as $a \to 0$, the integral goes of to $-\infty$ due to the same reasons. So, L takes all negative values as well. Thus, L is a continuous bijection.

Since L is a bijection from $\mathbb{R}_{>0} \to \mathbb{R}$, it has an inverse. We will first show that this inverse is continuous. To do this, we show that L is an open map, i.e., it sends open sets to open sets.

Given an open interval $(a, b), b > a > 0$, since L is increasing, this interval goes to $(L(a), L(b))$ in a bijective manner. So, L is an open map, which means that its inverse is continuous for given $\epsilon > 0$, for any x, the interval $(x - \epsilon, x + \epsilon)$, when modified to lie in $\mathbb{R}_{>0}$ goes to an interval, so contains some δ neighbourhood of $L(x)$.

Thus, L is a continuous bijection with a continuous inverse, that is L is a homeomorphism of $(0, \infty)$ with $\mathbb{R}$. Further, since L is differentiable, we will show that its inverse is differentiable as well.

Let us denote by f the inverse of L. We have shown that f is continuous, we want to find the derivative of f. Fix an $x \in \mathbb{R}$,

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{L(f(x+h)) - L(f(x))}$$

where we have used the property that $L \circ f = \mathrm{id}_{\mathbb{R}}$. Now as $h \to 0$, by continuity of $f, f(x+h) - f(x) \to 0$, so the limit above exists (by injectivity, numerator is never zero) and we have

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \frac{1}{L'(f(x))} = f(x)$$

because the derivative of L at x is $1/x$. Note that it is crucial that the derivative of L is non-zero, else f wouldn't be differentiable.

Thus we have arrived at a function f given by the inverse of L and the derivative of f is itself. Since $L(1) = 0$, we have $f(0) = 1$. That's all fine, but we have left the intuitive inverse hanging. We still don't know whether $\lim_{x\to\infty}(1+\frac{a}{x})^x$ exists. To show that this limit exists, not that one can take the log, L and look at $\lim_{x\to\infty} xL(1+\frac{a}{x})$. We use the linear approximation,

$$L(1+h) = L(1) + L'(1)h + R$$

where $\frac{R}{h} \to 0$ as $h \to 0$. Using this, we write

$$xL\left(1 + \frac{a}{x}\right) = x\frac{a}{x} + xR = a + xR$$

where $R\frac{x}{a} \to 0$ as $x \to \infty$. Thus $xR \to 0$ as $x \to \infty$ and hence

$$\lim_{x\to\infty} L\left(\left(1 + \frac{a}{x}\right)^x\right) = a.$$

Since f is continuous, we can apply f and we see that

$$\lim_{x\to\infty} \left(1 + \frac{a}{x}\right)^x = f(a).$$

As you can see, the intuitive inverse, is actually the inverse of L, and we have arrived at the exponential function. Moreover, we have proved that it is the inverse of L and is its own derivative.

Furthermore, $f(x+y) = f(x)f(y)$ because $L(f(x+y)) = L(f(x)) + L(f(y)) = L(f(x)f(y))$ and this points towards the exponential nature of f, similar to L being logartihm of x with respect to some basis.

In fact, we have $L(f(1)^x) = xL(f(1)) = x \implies f(1)^x = f(x)$. Thus, f is actually exponentiation with $e = f(1)$ as the base. Then, it is immediate that L is logarithm to the base e.

There are many ways to reach the number e, one common way is through power series expansion, the other is through differential equations. What one does is to find a function f with $f(xy) = f(x) + f(y)$ and/or $f'(x) = f(x)$ satisfying $f(0) = 1$. But, at the time of writing this (today is 28th October 2019), I haven't explored differential equations and functional analysis enough to say anything else on this. Another way to do this is to use the binomial expansion (of the limit definition of e), and observe that the denominators cancel out. But his approach requires an interchange of limits (because the binomial expansion is an infinite sum) and I thought the way I have proved it is much neater. Besides, the conditions for interchanging limits are too many, that I didn't bother to look at them.

17 Power series at last

So, we have finally arrived at the magical functions $\ln x$ and e^x. Both of these functions are infinitely differentiable, i.e., they are C^∞. It is natural to look at their power series. Of the two, e^x is particularly well behaved.

We expand e^x about 0. We have

$$e^x = \frac{1}{0!} + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + e^c \frac{x^n}{n!}$$

for some c between x and 0. Now we need to show that the remainder term goes to zero, then we know that the power series actually converges to e^x. To do this, we note that

$$e^c \leq e^{|x|}$$

for any c between x and 0 because e^x is an increasing function. So, we have an upper bound on e^c, all we now need to do is to show that the rest of the remainder term goes to zero (that is one way to proceed that works for this particular function; in general the situation may not be so nice, either the derivative term alone goes to zero or neither of them go to zero, but together they could).

There is an integer M such that $|x| < M$. So, $x^n < M^n$ for $n \geq 1$ and

$$\frac{x^{M+n}}{(M+n)!} < \frac{M^{M+n}}{(M+n)!} = \frac{M^n}{(M+n)(M+n-1)\ldots(M+1)} \frac{M^M}{M!} < \left(\frac{M}{M+1}\right)^n \frac{M^M}{M!}.$$

Since $\frac{M}{M+1} < 1$, it is clear that the remainder term does go to zero. Thus we have for all $x \in \mathbb{R}$

$$e^x = \frac{1}{0!} + \frac{x}{1!} + \frac{x^2}{2!} + \cdots = \sum_{n \geq 0} \frac{x^n}{n!}.$$

Next, for log one can do a similar expansion, but I leave it to the reader, this essay turned out to be a lot longer than I expected. Besides, log's expansion is not on the whole of $\mathbb{R}$. What one has to do is to expand $\log(1+x)$ about zero, and the radius of convergence turns out to be 1.

The radius of convergence of a power series is a number that determines the region where the power series converges absolutely, elsewhere it may converge conditionally or may not converge at all. There is a theorem that says that every power series has a radius of convergence (it may be zero) and there are methods to compute this radius using the coefficients of the series. As a consequence, the region where the series converges is always a disc (in $\mathbb{R}$ this is an open interval, in $\mathbb{C}$ it would be an open disc). Furthermore, treating the partial sums as a sequence of functions, one can show that the power series converges uniformly on compact regions contained in the region of convergence. Also, there are results that say when one can add, multiply, compose two series or differentiate/integrate a series term by term (this again deals with uniform convergence as proved above).

18 In the context of complex numbers

A natural extension of exponentiation is to the complex plane. Exponentiating to integral powers is just the usual multiplication and taking inverses. The problem lies in taking rational or irrational powers. Since the complex plane is algebraically closed, there is now way to assign a unique rational power of a complex number. So, it is futile to try to extend these notions the way we did on the real line. What else can we try? We can try to extend the power series. May be we can have notions that when restricted to the real line give us back what we started with. As I mentioned earlier, e^x is particularly well behaved. Since its power series has infinite radius of convergence, we can blindly extend it to the complex plane. So we define

$$e^z = \sum_{n\geq 0} \frac{z^n}{n!}, \forall z \in \mathbb{C}.$$

This is a definition. Is it actually e raised to z, well not really. If we want to define such a thing, let us start with $z = a + ib$. Now, if we want to make sense of exponentiation, we would like $e^z = e^a.e^{ib}$. The first term is fine, what about the second one? We must have a sensible definition of e^i. We cannot force this to be anything, we just have to look at what the series tells us.

So, if we plug in ib into the power series above, we get a bunch of terms. Now Euler collected the real and imaginary parts and arrived at a result. But we proceed with caution, can we just collect real and imaginary parts into two separate sums? Intuitively, it seems obvious, because the real and imaginary parts don't interact. But there is another problem, the definition of e^{ib} would be the limit of the partial sums, how do we know that this limit is the same as the sum of limits of two separate partial sums? One can prove that under absolute convergence, we can partition the original sum into finitely many independent parts (here it would be the even and odd terms). Using that result, this manipulation is valid.

If one does separate the real and imaginary parts, then one arrives at

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

I have changed b to θ for aesthetic reasons. Some people evaluate this at $\theta = \pi$ and call it the most beautiful formula in mathematics, but that is stupid. To arrive at this, one needs to first find the series expansions of trignometric functions, to do which one needs a definition of these trignometric functions, and show that they are differentiable and so on. Having found the expansion, one has to show that the radius is infinite and then this equation has a proof. I am going to skip all that.

Now that we have e^z for any z, can we talk about log? Note that in the real case, we had a natural progression. We defined exponentiation, then tried to find a certain derivative. In doing so, we arrived at the logarithm which had some nice properties. We then found the inverse of the logarithm, just because we knew it existed. This inverse had the fantastic property of being its own derivative. But in the complex plane, we can't take the first step we took in the real case. Could we force a definition of logarithm like we did e^z? No, because log doesn't have a power series on all of $\mathbb{R}$, which can't be extended, although $\log(1 + x)$ has a power series with a radius 1 which can be extended.

Even though we defined log as the derivative of something, most of its properties were derived from the observation that log was the integral of $1/x$. This definition can be extended. So, how does one proceed? In $\mathbb{R}$, we took the integral from 1 to any positive x. Why positive? A negative x would require us to integrate through 0 where the function is not defined. But in the complex plane, we can go to the negative side by going around zero.

That's where our first problem comes. Is the integral path independent? The answer is no, and the result was proved by Cauchy.

Of course, there is another way to define complex logarithm, and this is to brute force it. We know that log should be the inverse of e^z. Before we get into that, we verify a few things first.

A simple calculation gives us

$$e^{z_1+z_2} = e^{z_1}e^{z_2}, z_1, z_2 \in \mathbb{C}.$$

Further, using Euler's formula, every $z \in \mathbb{C}$ can be written as $z = re^{i\theta}$, which is just a fancy way of writing z in polar coordinates. So, if we want a w such that $e^w = z$, then if $w = a + ib$, we have

$$e^a e^{ib} = re^{i\theta}.$$

Taking complex modulus and using logarithm, we have $a = \ln r$. Cancelling terms gives us $e^{i(b-\theta)} = 1$. Finally using Euler's formula, we see that $b - \theta = 2n\pi$ for some integer n. This is the "brute force" method I mentioned above. However, this definition is not unique. Earlier, I mentioned that we could try to define log by taking path integrals of $1/z$. It turns out that the n above depends on how many times your path winds around 0.

This then leads us to the theory of holomorphic functions and complex logarithms.

Suppose we fix some n, say 0 and avoid the negative real line (else there are continuity issues). Then we have a definition of log, and it is now clear that we can take arbitrary powers. Given $a, b \in \mathbb{C}$, to compute a^b, just take the logarithm of a, multiply by b and then exponentiate. This is a sensible way to define exponentiation, and one can check that the usual rules hold.

19 Complex binomial theorem

Do we have a binomial expansion for complex numbers? Just as we did with the series for e^x, we can blindly extend the binomial series of $(1+x)^r$ to the complex plane for real r. We can define the binomial coefficients for complex numbers as we did for arbitrary real numbers

$$\binom{z}{n} = \frac{z(z-1)\dots(z-n+1)}{n!}.$$

Having defined the binomial coefficient, we consider the power series $\sum_{n\geq 0}\binom{b}{n}a^n$ and find its radius of convergence. For $b \notin \mathbb{N}$, the successive ratio of absolute coefficients is $|\frac{b-n}{n+1}|$ which converges to 1, so the radius of convergence of this power series is 1. Now, does this power series give $(1+a)^b$?

First we need to take the log of $1+a$. If we look at the Taylor series of $\ln(1+x)$, we have

$$\ln(1+x) = \frac{x}{1} - \frac{x^2}{2} + \dots + (-1)^{n-1}\frac{c^n}{n}$$

for some c between x and 0. It is clear that the remainder term goes to zero, and that the series $\frac{x}{1} - \frac{x^2}{2} + \dots$ converges to $\ln(1+x)$ with radius of convergence 1.

Suppose we take the same series for $1+a$ where a is a complex number, $|a| < 1$. We know that the series converges. Next we multiply this by b and exponentiate. We have

$$ba - \frac{ba^2}{2} + \frac{ba^3}{3} - \dots$$

The sequence we get when we do e^z where z is the partial sums is

$$1 + ba + \frac{b^2a^2}{2!} + \dots, 1 + \left(ba - \frac{ba^2}{2}\right) + \dots$$

and so on. A quick calculation shows that this sequence converges to $\sum_{n\geq 0}\binom{b}{n}a^n$. Note that e^z is a continuous function, so the limit of this sequence is what we get when we do e^z where z is given by the sequence above. In particular, setting $b = 1$ gives the limit as $1+a$. Therefore, the series of $\ln(1+x)$ we extended to the complex plane does give a series representation of some logarithm.

Thus, we have shown that the series for the binomial expansion extended to the complex plane is equivalent to the original definition of exponentiation (take log, multiply and then exponentiate). This proves the binomial expansion for the complex plane.

20 Conclusion

This essay started out with defining exponentiation in the context of real numbers. We took a real number, defined how to take nth powers, then rational and then using limits, irrational powers. Our domain was limited to the positive real numbers. Then we treated these as functions and studied the properties of these functions. We analysed the continuity and differentiability. On the way, we developed the binomial series and Taylor's theorems with many forms for the remainder term. Having done that, we looked at the function a^x for a fixed a. Continuity was established and to find its derivative, we needed the logarithm. From here, we went on to study logarithms. One could develop power series for a^x, but it seems unnecessary. Logarithms were shown to exist and turned out to be paritcular integrals. Having discovered that, we established the function e^x. Finally, we derived the power series of e^x.

Then we moved onto the complex plane. Here we have problems regarding rational powers and so on. The way around it is to define e^z, and logarithm. But logarithms are not uniquely defined. In order to exponentiate, we take the logarithm, multiply by the index, and then exponentiate again. Of course, this is not unique. Finally, we established a binomial expansion.

Now, one could have proceeded differently. One can, even in the real case, start with the power series for e^x. Show that this function is defined every where and is actually a number e raised to the power of x, in whatever way that makes sense and satisfies whatever properties we expect it to have. Then, define logarithm as the inverse (in the case of real numbers, this is well defined) and proceed to exponentiate by using logarithms. This is a valid development of the subject, except that I feel it strips out the essence of these functions. Why does it work? How does one come up with these definitions? Why a power series?

I fell that the subject has been developed sufficiently, and that I have collected most of my thoughts on this matter in this essay.

References

[1] Examples for uniform convergence and derivatives were taken from
https://www.math.ucla.edu/ tao/resource/general/131bh.1.03s/week45.pdf

[2] *Introduction to Real analysis* Bartle and Sherbert

[3] Some history on the binomial expansion was taken from
https://proofwiki.org/wiki/Binomial_Theorem/General_Binomial_Theorem

[4] The algebraic proof was taken from Paramanand Singh's answer here
https://math.stackexchange.com/questions/666796/elementary-proof-that-a-n-to-a-implies-a-nr-to-ar-for-r-in-mathbbq/667040#667040

[5] Method of exhaustion
https://en.wikipedia.org/wiki/Method_of_exhaustion

[6] Cavalieri's principle
https://en.wikipedia.org/wiki/Cavalieri%27s_principle

[7] Quadrature
https://en.wikipedia.org/wiki/Quadrature_(mathematics)

[8] Prosthaphaeresis
https://en.wikipedia.org/wiki/Prosthaphaeresis

[9] History of logarithms
https://en.wikipedia.org/wiki/History_of_logarithms
https://hsm.stackexchange.com/questions/635/how-did-napier-come-to-invent-logarithms
https://www.thocp.net/reference/sciences/mathematics/logarithm_hist.htm

[10] Adequality
https://en.wikipedia.org/wiki/Adequality

[11] Some history on calculus
http://www.math.wpi.edu/IQP/BVCalcHist/calctoc.html
http://www.unclebobpuzzles.com/Permasite/UB&AC/chapter3part2.htm

[12] All the different ways e^x is defined and how existence of limits is shown
https://en.wikipedia.org/wiki/Characterizations_of_the_exponential_function
https://math.stackexchange.com/questions/389793/what-is-the-most-elementary-proof-that-lim-n-to-infty-11-nn-exists
https://math.stackexchange.com/questions/359023/using-the-limit-definition-to-find-the-derivative-of-ex