The logarithm and exponential functions

See also: Lang chapter 13.

Review

Functions

To understand the relationship between exponentials and logarithms, we need to recall a few general facts about functions.

- graphs: a function has a graph; a curve in the plane is the graph of a funciton if and only if it passes the "vertical line test"
- one-to-one: a fuction is one-to-one if $f(a) = f(b)$ occurs *only* when $a = b$; a function is one-to-one if and only if its graph passes the "horizontal line test"
- Inverses: Some functions have inverses which "undo" the function. Sometimes an inverse is only "one-sided"
- Graphs and inverse functions: If f has an inverse, the graph of the inverse is obtained by reflecting the graph of f across the line $y = x$. Note that this swaps horizontal and vertical lines.
- A function has an inverse if and only if it is one-to-one. Functions which are not one-to-one can be restricted to regions where they are, and given inverses there. Different restrictions give different inverses.

Exponent rules

For any real (or complex) number a, and any natural number $n > 0$, we define $a^n = a \cdot a \cdots a$ (multiplied n times). For $n = 0$, we define $a^0 = 1$. This operation has the property that

$$a^m \cdot a^n = a^{m+n}$$

for any natural numbers m and n.

Now just as we extend the natural numbers, $\mathbb{N}$, using arithmetic rules to discover the integers, $\mathbb{Z}$, and then the rational numbers, $\mathbb{Q}$, we can extend the meaning of exponents to understand the meaning of a^q for every rational number q.

Exercise 1 *Use the exponent rule and the axioms for additive inverses to prove that, for $n > 0$, the exponent a^{-n} must mean $\frac{1}{a^n}$.*

Exercise 2 *Use the exponent rule and the axioms for multiplicative inverses to prove that, for any integer n, the exponent $a^{1/n}$ must mean $\sqrt[n]{a}$.*

Note that these don't give us a conceptual meaning of exponents, beyond the exponential rule. Also, there isn't an arithmetic we can use to extend exponents to irrational real numbers. But there *is* a way to do it without arithmetic! It is based on the Least Upper Bound property of the real numbers, which implies that the real numbers are precisely those which can be reached as the limit of a sequence of rational numbers. So to understand a^x for irrational numbers x, we use rational approximations (e.g. decimal approximations, or others). Equivalently—and importantly—you could think of graphing a^x for just the *rational* numbers x, and then complete the function a^x by filling in the holes in the graph!

For a conceptual meaning of exponentials, we need a little more depth of understanding from the sections below.

The exponential functions

- definition of $\exp_a(r)$ for rational r
- properties of exp
- example: doubling

The logarithm functions

- inverse to exponentials
- graph; properties of logarithms (the logarithm rule)

The natural logarithm: The area under the curve $1/x$

In this section, we describe a very clever observation: The area under the curve $y = \frac{1}{x}$ satisfies the logarithm rule! Here's what I mean:

Definition 1. *For each real number $t > 0$, let $A(t)$ be the area under the graph of $y = \frac{1}{x}$ from $x = 1$ to $x = t$. (If $0 < t < 1$, count this as negative area.)*

Theorem 1. *For any two real numbers $s, t > 0$, we have*

$$A(st) = A(s) + A(t).$$

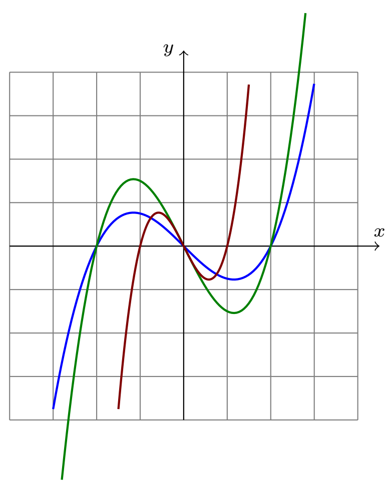
If you know about definite integrals, this theorem can be proved easily using a change of variables, letting $u = st$, so $du = sdt$. But this change of variables obscures a very simple geometric argument, which I'll describe now.

This becomes slightly easier to discuss if we expand the definition of A so that $A(a, b)$ denotes the area under $y = 1/x$ between $x = a$ and $x = b$ (positive if $a < b$, negative if $b < a$, and zero if $a = b$). Then our earlier definition is $A(t) = A(1, t)$.

Now it's clear, because area is additive, that $A(1, st) = A(1, s) + A(s, st)$. So to prove the theorem we just need to explain why $A(s, st)$ should be the same number as $A(1, t)$. For example, why should $A(3, 6) = A(1, 2)$?

As we will see, this all comes down to the very simple fact that $\frac{s}{st} = \frac{1}{t}$! How? First, recall that scaling a shape in the vertical direction multiplies its area by the scale factor. Same for scaling in the horizontal direction. And to connect with graphs, note that multiplying a function by a number scales its graph. Also, multiplying the *input* of a function scales its graph.

In general, the graph of $s \cdot f(x)$ is scaled vertically by a factor of s, while the graph of $f(s \cdot x)$ is scaled horizontally by a factor of $\frac{1}{s}$. For example, the graphs of $f(x)$, $2f(x)$ and $f(2x)$ are shown below, where $f(x)$ is the cubic polynomial $x^3 - x$.



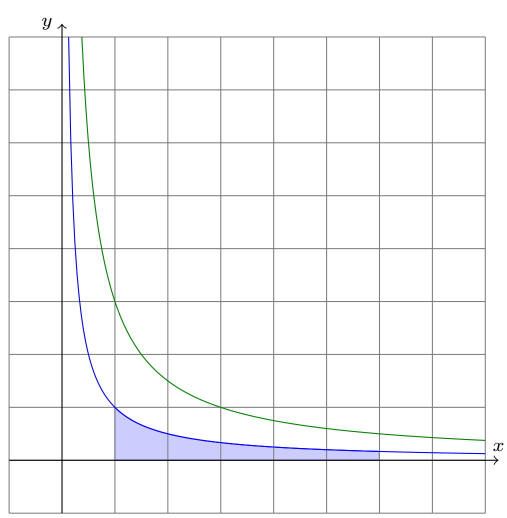
Graphs of $y = f(x)$ (blue); $y = 2f(x)$ (green); and $y = f(2x)$ (red).

Now what does this mean for our area function? Well, the area under $2f(x)$ on some interval $[a, b]$ will be 2 times the area under $f(x)$ (on that same interval). Likewise, the area under $f(2x)$ on $[\frac{a}{2}, \frac{b}{2}]$ will be $\frac{1}{2}$ times the area under $f(x)$ on the interval $[a, b]$. We have to change the interval because, in order to get $2x$ ranging over the same interval $[a, b]$, we have to restrict x to the interval $[\frac{a}{2}, \frac{b}{2}]$.

Both of these multiplications scale the graph, and hence the area, in one direction or the other (one vertical, one horizontal). Now here's a cool fact: this means that the area under $2f(2x)$, on the interval $[\frac{a}{2}, \frac{b}{2}]$ will be *exactly the same* as the area under $f(x)$ on the interval $[a, b]$. And of course this works more generally: for any $s > 0$, the area under $s\,f(sx)$ on the interval $[\frac{a}{s}, \frac{b}{s}]$ will be the same as the area under $f(x)$ on the interval $[a, b]$. Note that, in general, the function $s\,f(sx)$ is pretty different from $f(x)$ – it is stretched by a factor of s vertically, and compressed by a factor of $\frac{1}{s}$ horizontally. But these changes balance out to preserve area (from an original interval to a compressed one).

The previous paragraphs tell us that the area under $f(x)$ between s and st (that is, the interval $[s, st]$) will be the same as the area under $s\,f(sx)$ betweeen 1 and t (that is, the interval $[1, t]$)!! Now we come to the special fact about the function $f(x) = \frac{1}{x}$: for this function, $s\,f(sx)$ is exactly the same function as the original $f(x)$. And this tells us that the area $A(s, st)$ is exactly the same as the area $A(1, s)$!

Lastly, let's point out that the same discussion we've had would apply equally well to $f(x) = \frac{2}{x}$, or $f(x) = \frac{3}{4x}$, or any other constant times $\frac{1}{x}$. These give lots of different functions that all satisfy the same logarithm rule, and the most basic—most *natural*—is the one we get from $\frac{1}{x}$.



Graphs of $y = 1/x$ (blue) and $y = 3/x$ (green).
The area under $y = 1/x$ between 1 and 6 is shaded.

Definition 2. *The* natural logarithm*,* $\ln(x)$ *is defined to be the function* $A(x)$

above.

Now, of course, we have a question: what is the *base* of the natural logarithm? It is the number e such that $\ln(e) = 1$. Said differently, it is the number e such that the area under $\frac{1}{x}$ between 1 and e is precisely 1. This number is called *Euler's number*. Euler was one of the first mathematicians to understand the importance of this number, and used it some of his work in the 1720s. It was discovered about 50 earlier, by one of the Bernoulli brothers (Jacob) in a different context (studying compound interest).

Exercise 3 *Use the area definition of* $\ln$ *to prove that* $2 < e < 4$.

The natural exponential

- The natural exponential is the one which is inverse to the natural logarithm.
- A conceptual meaning of exponentials: functions whose rate of change is proportional to their values. (This can be said more clearly with derivatives.) The natural exponential is the one whose constant of proportionality is 1.