Summary of material on the natural logarithm and e, in three acts

Yvonne Lai

University of Michigan – Math 486

In writing these materials and creating problem sets 8-10, I have drawn from the following sources:

- Courant and Robbins. *What is Mathematics? An Elementary Approach to Ideas and Methods*. This book was first published in 1941 to explain some of the most beautiful ideas of mathematics. You may be interested in knowing that discussions with many scientists and mathematicians (including Niels Bohr!) influenced the material in this book.
- Rusczyk and Lehoczky. *The Art of Problem Solving, Volume 2*. The Art of Problem Solving publishes resources for improving proof and problem solving ability. Their books often contain slick explanations and proofs.
- Usiskin, Peressini, Marchisotto, and Stanley. *Mathematics for High School Teachers – An Advanced Perspective.* The required text for Math486.
- Connally, Hughes-Hallett, et al. *Functions Modeling Change: A Preparation for Calculus*. The textbook for Math105.
- Apostol. *Calculus, Vol. 1: One-Variable Calculus with an Introduction to Linear Algebra*. An introduction to the theoretical foundations of one-variable calculus, including integration, differentiation, and properties of functions.

Prologue: Injective and Invertible Functions

- A *function* is a mapping from inputs to outputs, where every input is assigned exactly one output.
- A function is *injective* if for every output, there is only one input. In other words, $f : \mathit{Domain} \to \mathit{Range}$ is injective if for $f(x_1) = f(x_2)$ implies $x_1 = x_2$ for all x_1, x_2 in the domain of f. Injective functions are also called *one-to-one* functions.
- Suppose $f : \mathit{Domain} \to \mathit{Range}$ is a function. Then $f^{-1} : \mathit{Range} \to \mathit{Domain}$ is an *inverse function* of f if $f^{-1} \circ f(x) = x$ for all x in the domain of f.
- A function f is *invertible* if it has an inverse function of f.

You may notice the phrase "*an* inverse function", which implicitly suggests that we cannot a priori rule out the possibility of more than one inverse function. To establish that a function only has one inverse requires a proof:

Proposition. Suppose $f : \mathit{Domain} \to \mathit{Range}$ is an invertible function. Then f has a unique inverse function.

In the proof of the proposition, we will use D to denote the input space of f and R to denote the output space of f.

Proof. Suppose $g : R \to D$ and $h : R \to D$ are inverse functions of f, so $g \circ f(x) = x$ and $h \circ f(x) = x$ for all x in the domain of f. We show that g and h must be the same function: that is, $g(y) = h(y)$ for all elements y in R, the domain of g and h.

Suppose y is in R. Then y is in the range of f, so there is an input $x_0 \in D$ so that $f(x_0) = y$, so $g(y) = g \circ f(x_0) = x_0$. Similarly, $h(y) = h \circ f(x_0) = x_0$. Hence $g(y) = h(y)$. We have not made any assumptions about y aside from being an element of R, so $g(y) = h(y)$ holds for all elements y in R, the domain of g and h. Therefore g and h are the same function, and invertible functions have unique inverses. $\square$

Another thing that we want to know for sure: if f^{-1} is the inverse of f, does that mean that f is the inverse of f^{-1}? According to our definition, this means that we need to know that $f \circ f^{-1}(y) = y$ for all y in the domain of f^{-1}.

Proposition 1. *Let f be an invertible function. Then f is the inverse of f^{-1}.*

Proof. We must show $f \circ f^{-1}(y) = y$ for all y in the domain of f^{-1}. If y is in the domain of f^{-1}, then it is in the range of f, so there exists an element x in the domain of f such that $f(x) = y$. Thus $f \circ f^{-1}(y) = f \circ f^{-1} \circ f(x) = f(x) = y$. $\square$

A property of invertible functions we discussed in class is the passing of the *Horizontal Line Test*. The horizontal line test says that a function is invertible if and only if for every output, there is one input giving that output. One result of our discussion was that **injectivity implies invertibility**.

Other useful properties involving invertible functions include:

- decreasing implies invertibility (i.e., if f is a decreasing function, then f is invertible.)
- increasing implies invertibility
- if f is an invertible, differentiable function, then

$$\frac{df^{-1}}{dy}(y) = \frac{1}{\frac{df}{dx}(x)}.$$

  One can prove this property by applying chain rule for differentiation on the definition of inverse function: $f^{-1} \circ f(x) = x$.
- given an invertible function f, the graph $y = f^{-1}(x)$ can be obtained by reflecting the graph $y = f(x)$ over the line $y = x$.
  To prove this property, we first showed that if (x_0, y_0) is a coordinate in the plane, then (y_0, x_0) is the reflection of (x_0, y_0) over the line $y = x$.
  Now suppose that (x_0, y_0) is a point in the graph $y = f(x)$. This means that $y_0 = f(x_0)$ by definition of the graph. When this point is reflected over the line $y = x$, we obtain the point (y_0, x_0), which satisfies the equation $y = f^{-1}(x)$ because $f^{-1}(y_0) = f^{-1} \circ f(x_0) = x_0$. So when the graph of $y = f(x)$ is reflected over the line $y = x$, we obtain points in the graph $y = f^{-1}(x)$. Since the domain of f^{-1} is the range of f, we can obtain all points in $y = f^{-1}(x)$ in this way. $\square$

Inverse Trigonometric Functions

Sometimes, we will want to be able to force an inverse function for a function that shouldn't be invertible. The most famous examples of this are the trigonometric functions. Even though sin and cos are not invertible functions, we still refer to functions that we call $\sin^{-1}$ and $\cos^{-1}$. (These functions are also known as arcsin and arccos.)

We define $\sin^{-1}$ and $\cos^{-1}$ is by restricting the domain of sin and cos.

function	usual domain	range
sin	$\mathbb{R}$	$[-1, 1]$
cos	$\mathbb{R}$	$[-1, 1]$
function	restricted domain for obtaining inverse	range
sin	$[-\frac{\pi}{2}, \frac{\pi}{2}]$	$[-1, 1]$
cos	$[0, \pi]$	$[-1, 1]$
inverse function	domain of inverse function	range of inverse function
$\sin^{-1}$	$[-1, 1]$	$[-\frac{\pi}{2}, \frac{\pi}{2}]$
$\cos^{-1}$	$[-1, 1]$	$[0, \pi]$

So $\sin^{-1}$ is defined as the inverse of the function $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$ and $\cos^{-1}$ is the inverse of the function $\cos : [0, \pi] \to [-1, 1]$.

Act I: The Natural Logarithm and e^x as the inverse of $\ln$

One of the first results we learn in calculus is $\frac{d}{dx}(x^n) = nx^{n-1}$. We're excited to see that it works for whole numbers such as 2, 3, and 100, negative numbers and even irrationals. The only number that it doesn't work for is $n = -1$.

The Fundamental Theorem of Calculus comes to the rescue. We define the function $\ln(x)$ as

$$\ln(x) = \int_1^x \frac{1}{u}\,du.$$

This function satisfies $\frac{d}{dx}(\ln(x)) = \frac{1}{x}$.

The value of $\ln(x)$ is positive when we integrate to the right, and $\ln(x)$ is negative when we integrate to the left. For example:

$$\begin{aligned} \ln(0.5) &= \int_1^{0.5} \frac{1}{u}\,du \\ &= -\int_{0.5}^{1} \frac{1}{u}\,du. \end{aligned}$$

We can think of the integral from 1 to the input of ln as measuring the area under the curve $y = \frac{1}{x}$ from 1 to where we evaluate ln, where we negate the area when the input is to the left of 1.

This function is known as the *natural logarithm function*. (The letters are backwards because of its French origins: "logarithm naturel".) As we will learn, the natural logarithm and the function e^x are inverses of each other.

We define the domain of ln to be $\mathbb{R}_{>0}$.

Beautiful properties of the logarithm fall out of its definition.

- $\ln(1) = 0$, as the area under the curve $y = \frac{1}{x}$ between 1 and itself is 0.
- Given $a, x \in \mathbb{R}$, we have $\ln(ax) = \ln(a) + \ln(x)$. To show this property, observe that $\frac{d}{dx}(\ln(ax)) = \frac{1}{ax} \cdot a$ by the chain rule. So $\frac{d}{dx}(\ln(ax)) = \frac{1}{x} = \frac{d}{dx}(\ln(x))$. When two functions have the same derivative, they must differ by a constant: there is a constant C such that

  $$\ln(ax) = \ln(x) + C$$

  for all $x \in \mathbb{R}$. Because the above equality holds for all x, and C is a constant (so it does not depend on x), we know that

  $$\ln(a \cdot 1) = \ln(1) + C.$$
  $$\ln(a) = C.$$

  We have shown that $\ln(ax) = \ln(x) + \ln(a)$ as desired.
- ln is an increasing function. To explain why this property is true, observe that its derivative is $\frac{1}{x}$, which is positive on the domain of $\ln(x)$.
- ln is injective. This follows from the fact that ln is increasing.
- $\ln(x^n) = n\ln(x)$. We can show this in a similar fashion to the identity $\ln(ax) = \ln(a) + \ln(x)$.

Now that we know that ln is an injective function, we know that there is only one input for which the output is 1. We call this special input e:

$$\ln(e) = 1.$$

Miraculously, identifying this number allows us to identify the inverse function of ln! It is e^x.

Proposition 2. *Let $f(x) = e^x$. Then $f^{-1}(x) = \ln(x)$.*

Proof. We must show that $f^{-1} \circ f(x) = x$ for all x in the domain of f, which is $\mathbb{R}$ in this case. Observe that $\ln(e^x) = x\ln(e) = x$. $\square$

The function e^x has a number of exceedingly pretty properties. One of them, which is very useful for engineering contexts, is:

Proposition 3. *A function f is its own derivative if and only if it has the form $f(x) = ae^x$, where a is a constant.*

This is incredible! It means that the only way that f can satisfy the equation $f = \frac{df}{dx}$ is if f is e^x or a multiple of e^x.

Proof. [$\Longleftarrow$] First we show that if $f(x) = ae^x$, then $\frac{df}{dx} = ae^x$. Recall that

$$\ln(e^x) = x.$$

Then taking the derivative of both sides, we obtain

$$\begin{aligned}
\frac{d}{dx}\ln(e^x) &= \frac{d}{dx}(x) \\
\frac{1}{e^x}\cdot\frac{d}{dx}(e^x) &= 1.
\end{aligned}$$

Hence $\frac{d}{dx}(e^x) = e^x$. It follows that $\frac{d}{dx}(ae^x) = ae^x$ for a constant a.
[$\Longrightarrow$] Now we show that if $\frac{d}{dx}f(x) = f(x)$, then $f(x) = ae^x$ for some constant a.

If $\frac{d}{dx}f(x) = f(x)$, then

$$\begin{aligned}
\frac{d}{dx}f(x) &= f(x) \\
\frac{1}{f(x)}\,\frac{df(x)}{dx} &= 1 \\
\int \frac{1}{f(x)}\,\frac{df(x)}{dx}\,dx &= \int 1\,dx \\
\int \frac{1}{f(x)}\,\frac{df(x)}{\cancel{dx}}\,\cancel{dx} &= \int 1\,dx \\
\int \frac{1}{f(x)}\,df(x) &= \int 1\,dx.
\end{aligned}$$

Integrating tells you that f satisfies

$$\ln(f(x)) = x + C.$$

To complete the proof (this was part of your homework), make both sides exponents of e and try to get the desired conclusion (there exists a constant a such that $f(x) = ae^x$). $\square$

Act II: The Binomial Theorem and Pascal's Triangle

Suppose you are teaching AP Calculus. You assign your students n problems for homework, but tell them that they only need to turn in k of them, where $0 \le k \le n$. In this possibly unrealistic scenario, you also make an answer key for each of the possible subsets of problems that your students turn in. How many answer keys do you make up?

We use the notation $\binom{n}{k}$ to denote the number of answer keys, which is

$$\frac{n!}{(n-k)!k!}.$$

Here $n!$ means $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$. For example, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$, and $\binom{5}{2} = \frac{5\cdot4\cdot3\cdot2\cdot1}{(3\cdot2\cdot1)(2\cdot1)} = 10$. We read the notation $\binom{n}{k}$ as "n choose k."

The fundamental observation behind the quantity $\binom{n}{k}$ is the following:

The number of ways to line up n objects in a line is $n!$.

For example, there are $3! = 6$ ways to line up the letters A, B, C: there are **three** letters that could be first in line. After choosing the first in line, there are **two** letters left to choose from for the second in line. After selecting the first two letters, there is only **one** letter remaining. In total, there are $3 \cdot 2 \cdot 1 = 6$ ways to line them up.

Using similar reasoning, we can deduce:

There are $n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}$ ways to choose and line up k objects out of n total objects.

Back to the AP Calculus scenario – suppose there are $n = 20$ possible homework problems and you would like to select and sequence 3 for your students to do, then there are $20 \cdot 19 \cdot 18$ ways: you have 20 choices for the first problem. Once you have chosen the first problem, you have 19 problems left to choose from for the second problem. Then you have 18 problems left to choose from for the third problem. In total, there are $20 \cdot 19 \cdot 18 = \frac{20!}{17!}$ ways to select and sequence problems.

Finally, let us examine the original problem. We explain why:

There are $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ ways to choose k objects out of n total objects.

If there are $n = 20$ possible homework problems (labelled A, B, C, etc.), you only ask students to turn in any 3 of their choice, and now you need to make up the answer keys. If you make up an answer key for every possible order of every selection of 3 problems that your students turn in, then you would have to make up $20 \cdot 19 \cdot 18$ answer keys. However, you should only have to make up one answer key for the students who turn in the problems A, B, C regardless of what order they write up those problems in. Since there are 6 ways that they could order the problems in their write-up, for every 6 possibilities, you only need to do one answer key. So$20 \cdot 19 \cdot 18$ is 6 times more answer keys than you should be writing, meaning you need to write $\frac{20 \cdot 19 \cdot 18}{6} = \binom{20}{3}$ answer keys in total.

These numbers, known as binomial coefficients, also show up in the polynomial $(x+y)^n$; the **Binomial Theorem** states:

The coefficients of the polynomial $(x+y)^n$ are given by

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k.$$

For example:

$$\begin{aligned} (x+y)^3 &= \binom{3}{0}x^3 + \binom{3}{1}x^2y + \binom{3}{2}xy^2 + \binom{3}{3} = x^3 + 3x^2y + 3xy^2 + y^3 \\ (x+y)^{20} &= \binom{20}{0}x^{20} + \binom{20}{1}x^{19}y + \binom{20}{2}x^{18}y^2 + \cdots + \binom{20}{18}x^2y^{18} + \binom{20}{19}xy^{19} + \binom{20}{20}y^{20}. \end{aligned}$$

A remarkable thing occurs when the coefficients of $(x+y)^n$ are arranged as follows:

$$\begin{array}{c} \binom{0}{0} \\ \binom{1}{0}\binom{1}{1} \\ \binom{2}{0}\binom{2}{1}\binom{2}{2} \\ \binom{3}{0}\binom{3}{1}\binom{3}{2}\binom{3}{3} \\ \cdots \end{array}$$

This triangle is known as *Pascal's Triangle*, and its amazing property is that an interior term is always the sum of the two terms directly above it. For example, $\binom{2}{1} = \binom{1}{0} + \binom{1}{1}$, $\binom{3}{1} = \binom{2}{0} + \binom{2}{1}$, $\binom{3}{2} = \binom{2}{1} + \binom{2}{2}$. The exterior terms are all 1. This means we can generate the next row with just addition, no multiplication!

In Act III, we will make use of the binomial theorem for the polynomial $(1+\frac{x}{n})^n$. The binomial theorem tells us that

$$\left(1+\frac{x}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{x}{n}\right)^k = \sum_{k=0}^{n} \frac{n!}{(n-k)!k!} \cdot \frac{x^k}{n^k} = \sum_{k=0}^{n} \frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot 1}{n^k} \cdot \frac{x^k}{k!}$$

Act III: e can also be defined as the constant for continuously compounded interest

A pot of gold contains 1 million grams of gold. The wizard who found it is about to cast a spell that doubles it each year, in other words, increases by it 100% each year. So t years afterwards, the pot of gold would hold

$$2^t = (1+1)^t \text{ million grams.}$$

A second wizard comes by and suggests that instead of increasing by 100% each year, the first wizard should increase the worth by 1/2 each half year, so in t years, the pot would hold

$$\left(1+\frac{1}{2}\right)^{2t} \text{ million grams.}$$

Before the first wizard can take this suggestion, a third wizard swings by and suggests increasing by 1/3 each third of a year, so in t years, the pot would hold

$$\left(1+\frac{1}{3}\right)^{3t} \text{ million grams.}$$

Quickly, a fourth wizard jumps in with the suggestion to have, in t years,

$$\left(1+\frac{1}{4}\right)^{4t} \text{ million grams.}$$

As a fifth wizard materializes, the original wizard exhorts, "Stop! I have a better suggestion than all of you!" The wizard leaps and begins to cast the spell that would make his pot of gold hold, in t years,

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{nt} \text{ million grams.}$$

The fifth wizard cuts the first wizard off, nearly knocking off both their caps. "How do you know that limit exists? If the limit is ∞, then the world will collapse! Even we are not powerful enough to sustain infinite amounts of gold!"

The fourth wizard sniffs, "And how do you know that even if the limit isn't ∞, that your idea is better than mine?"

The wizards are discussing here the notion of compounded interest: for example, an amount of money accrues interest rate of 100% compounded 4 times a year, if every quarter year, the money increases by 25% of its worth at that moment. The final suggestion of the original wizard is to produce *continuously compounded interest*.

For the sake of brevity, let $a_n = \left(1+\frac{1}{n}\right)^n$. The wizards lay out the following issues, which we will discuss:

- How do we know that $a_1 < a_2 < a_3 < a_4 < \ldots$?
- How do we know that $\lim_{n\to\infty}(1+\frac{1}{n})^n$ is a meaningful limit?

After we resolve these issues, we will show how the sequence $\{a_n\}$ brings us full circle, back to e. We will discuss:

- Why the mystery number $M = \lim_{n\to\infty} a_n$ satisfies the identity $M^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$.
- Why $\ln(M) = 1$, meaning that $e = \lim_{n\to\infty} a_n$.

(Stay tuned for the final chapter! To be handed out on Wednesday.)