e, *i*, *e*, *i*, oh!

Joshua E. Hill

Preface and Appendix by Ernest E. Hill

Draft 2006-09-18

Preface (stub)

Ernie's preface on the origin of his unpublished treatise on e.
Talk about unpublished treatise on e presented upon teaching calculus
Include it as reference (1994)
Appendix, with preface to appendix

Foreword

The purpose of this paper is to discuss the standard definitions, meanings, and selected applications for the fabulously useful numbers *e* and *i*. These numbers are used at almost all levels in collegiate mathematics, but there is seldom an attempt to unify their various apparently disparate definitions or impart a sense of fundamental meaning for these constants.

In my pursuit of these goals, I have embraced mathematical formalism where it seemed necessary for the development of a convincing definition. In some places we get into areas of mathematics which are seldom taught outside of a Mathematics curriculum. My apologies for the wading that is required in these areas, but it seemed a disservice to shield the reader from the core issues at hand. In all cases, proofs are present to show how different ideas are connected and are intended to aid in understanding. If you find that the proofs don't help you, feel free to skip them.

No attempt has been made to force this paper to progress in a way that parallels the historical development of these constants. I have made some attempt to occasionally pause and reflect on some of the historical aspects, but these are thoroughly covered elsewhere in detail (for example, see the book e: The story of a Number by Eli Maor and).

I have made an attempt to develop most of the material that is necessary to understand this material. I have attempted to minimize the explicit prerequisites: a college level calculus background is assumed, ideally with a good grounding in infinite series. It would be a self-aggrandizing statement (bordering on absurdity) to suggest that this paper is geared toward "a sufficiently motivated high school student with a calculus background". Certainly, I would not have been able to approach this material at that time in my education. An undergraduate math major who has had a few classes in Linear Algebra, Abstract Algebra, Real and Complex Analysis, Combinatorics and Number Theory would find most of this content review, though a few surprises may lurk in unexpected places.

I have attempted to develop applications from widely disparate areas of human endeavor. My goal in this is to show that constants *e* and *i* are used so commonly that the standard question "Where are these constants used?" would be more usefully phrased "Where are these constants **not** used?" The various applications are largely independent, so skipping around is encouraged.

In addition, *e* and *i* are so centrally located in mathematics that we can naturally explore several fundamental theorems along the way. These topics are so attractive that it seems a shame to pass near them without exploration.

If you have any suggestions for the prose or mathematical content of this essay, please e-mail them to me.

Joshua Hill <josh-eieioh@untruth.org>

Table of Contents

1 *e* in the Real Numbers

1.1 Definitions of *e* and the exponential function

Math is a branch of human endeavors touched by aesthetic concerns, so the selections of definitions are both a matter of utility as well as a matter of taste. As the constant *e* is of fundamental importance in many different branches of math, there are several (ultimately equivalent) definitions for this important constant. The two traditional starting points are the series definition of *e* (which we cover as definition 1e) and the integer based limit (which we cover as definition 1c). We will start with a more intuitive view and work to the more standard definitions.

We will examine some of the common ones.
Let's start with the definition that I think reflects the center of *e*:
Def 1: *e* is the positive constant such that the derivative of the exponential function $E(x)=e^x$ is itself. (i.e., $\frac{d}{dx}\left[e^x\right]=e^x$)
So, that's a great proscriptive statement, but is there such a number? Let's explore the characteristics that such a number would have, and see if the existence of such a number falls out anywhere.
The definition of the derivative is $f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$. We are looking for an exponential function, which is of the form $f(x)=a^x$ (where *a* is some positive constant), so

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}\frac{a^{x+h}-a^x}{h}=\lim_{h\to 0}\frac{a^x a^h-a^x}{h}=a^x\left[\lim_{h\to 0}\frac{\left(a^h-1\right)}{h}\right].$$

Now, note that $f'(0)=\lim_{h\to 0}\frac{f(0+h)-f(0)}{h}=\lim_{h\to 0}\frac{a^h-1}{h}$,
so $f'(x)=a^x f'(0)=f(x)f'(0)$. This implies that the derivative of any exponential function is the original function, with some scaling constant applied. Further, this scaling constant is the slope of the tangent line to the function at *x*=0.

In order to find our value *e*, we are interested in finding the value *a* for which the scaling value is 1, so that we'll have the happy case where $f(x)=f'(x)$. This brings us to the next conventional definition:

Def 1a: *e* is the constant value such that the slope of the tangent line to the function $f(x)=e^x$ at *x*=0 is 1. (i.e., $f'(0)=1$)

We are interested in finding the value such that this is a true statement. First, why should such a value exist? Figure 1 shows two exponential functions and their tangent lines at *x*=0: $f(x)=3^x$ (the green curve), the tangent line to $f(x)$ at *x*=0 (the green dashed line), $g(x) = 2^x$ (the blue curve), and the tangent line to $g(x)$ at x=0 (the blue dashed line). The reference curve of slope 1 at x=0 is the dashed red line. Given that any value for a can be chosen between 2 and 3, it makes sense that any slope between the two noted tangent lines should be obtainable.

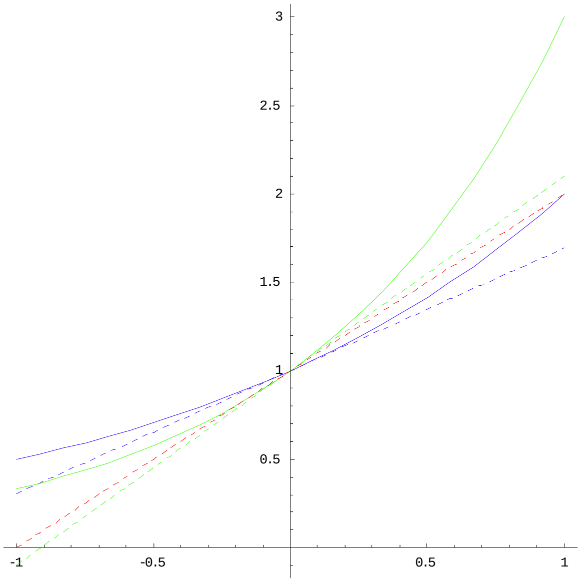


Figure 1

Sadly, there is a fairly significant difference between intuiting why something ought to be true, and actually proving that it is true (and even more sadly, many of the things that intuitively appear true end up being false).

Let's explore the fact that $f'(0) = 1$. First, let's suggestively re-title our constant to e.

$$f'(0) = \lim_{h\to 0} \frac{e^h - 1}{h} = 1 \text{, so}$$

$$\lim_{h\to 0} e^h - 1 = \lim_{h\to 0} h \Leftrightarrow \lim_{h\to 0} e^h = \lim_{h\to 0} 1 + h \Leftrightarrow \lim_{h\to 0} e = \lim_{h\to 0} (1+h)^{\frac{1}{h}} = e \,.$$

So, we have our third definition:

Def 1b: $e = \lim_{h\to 0} (1+h)^{\frac{1}{h}}$

Does this limit exist? Well, $f(h) = (1+h)^{\frac{1}{h}}$ is certainly going to be continuous for values of h in the domain $(-1,0)\cup(0,1)$, so we're only concerned with the behavior of the left and right limit of $f(h) = (1+h)^{\frac{1}{h}}$ as h goes to 0; in particular, we want the left hand and the right hand limit to be equal. Again, let's wave our hands, try to explain why it ought to work out, and then delay demonstrating that it does work out for just a bit longer.

Figure 2 is a graph of the function $f\left(h\right)=\left(1+h\right)^{\frac{1}{h}}$ near 0.

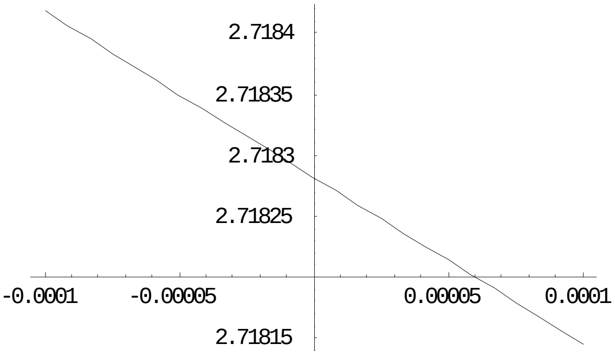


Figure 2

$f\left(h\right)$ behaves as if the function appears to approach the same value, both from the left and from the right of zero, so it would appear that this limit does exist.

Now, we can note that as *h* goes to zero, $\frac{1}{h}$ goes to infinity, so we can pose a related limit over the integers:

Def 1c: $e=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n}$

Again, we'll delay discussion on whether this limit exists for just a bit longer, other than to note that this is closely related to the definition 1b. If the limit in definition 1b exists, its right hand limit equals its left hand limit. Definition 1c is effectively one sequence that explores the right hand limit of 1b.

The corresponding left hand limit would look like this:

Def 1d: $e=\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{-n}$

In addition, because we know that the function $f\left(h\right)=\left(1+h\right)^{\frac{1}{h}}$ is continuous in the domain $\left(-1,0\right)\cup\left(0,1\right)$, we also know that if the limits in 1c and 1d both exist and are equal, then the limit in 1b exists.

We limited 1c to the integers, so we can look at this using the (standard, integer based) binomial theorem:

$\left(x+y\right)^{n}=\sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^{k}$, with $\binom{n}{k}=\frac{n^{\underline{k}}}{k!}$ where $n^{\underline{k}}$ is called the falling factorial, defined as $n^{\underline{k}}=\left(n\right)\left(n-1\right)\left(n-2\right)...\left(n-k+1\right)$ (this is equivalent to the more common description

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ when n is a positive integer, but we're going to find the first form more useful).

In this case, we will revisit definition 1c:

$$\left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n}\binom{n}{k}\frac{1}{n^k} = \sum_{k=0}^{n}\frac{n^{\underline{k}}}{k!n^k} = \sum_{k=0}^{n}\left[\frac{(n)(n-1)(n-2)...(n-k+1)}{n^k}\left(\frac{1}{k!}\right)\right]$$

Now, note that the degree (in n) of the numerator is k, as is the degree of the denominator. The coefficient of the n^k term is 1 in both the numerator and denominator, thus $\lim_{n\to\infty}\frac{(n)(n-1)(n-2)...(n-k+1)}{n^k} = 1$. So,

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = \lim_{n\to\infty}\sum_{k=0}^{n}\binom{n}{k}\frac{1}{n^k} = \lim_{n\to\infty}\sum_{k=0}^{n}\left[\frac{(n)(n-1)(n-2)...(n-k+1)}{n^k}\left(\frac{1}{k!}\right)\right] = \sum_{k=0}^{\infty}\frac{1}{k!}$$

Def 1e: $e = \sum_{k=0}^{\infty}\frac{1}{k!}$

It may not look like it, but in some sense this is the nicest definition we've hit so far. First, we can actually directly verify that this converges to an actual value, so we can for the first time authoritatively say that *e* exists!

As you may remember, $3 = 1 + \sum_{k=0}^{\infty}\frac{1}{2^k}$. Expanding the two series that we're looking at (which are both positive series), we see that each term of our *e* series is less than or equal to our 3 series:

$$e = \frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\frac{1}{4!}+...$$

$$3 = 1+\frac{1}{2^0}+\frac{1}{2^1}+\frac{1}{2^2}+\frac{1}{2^3}+...$$

The $k!$ term in the *e* series grows much faster than the 2^k term in the 3 series, and our 3 series converges, so the comparison test implies that our *e* series converges as well.

How quickly does it converge? Well, there are better estimates, but the obvious one from our development thus far is to note that all the terms after the *k*th term of *e* sum to a value less than $\frac{1}{2^k}$, so we can bound the tail of the series by $\frac{1}{2^k}$. That's pretty quick convergence, but we can do still better.

$$e = \sum_{k=0}^{n}\frac{1}{k!}+\frac{1}{(n+1)!}+\frac{1}{(n+1)!(n+2)}+\frac{1}{(n+1)!(n+2)(n+3)}+... <$$

$$\sum_{k=0}^{n}\frac{1}{k!}+\frac{1}{(n+1)!}+\frac{1}{(n+1)!(n+1)}+\frac{1}{(n+1)!(n+1)^2}+...$$

The error term here is a geometric series which converges to a known value, so

$$e=\sum_{k=0}^{n}\frac{1}{k!}+\sum_{k=n+1}^{\infty}\frac{1}{k!}<\sum_{k=0}^{n}\frac{1}{k!}+\frac{1}{(n+1)!}\sum_{k=0}^{\infty}\frac{1}{(n+1)^{k}}=\sum_{k=0}^{n}\frac{1}{k!}+\frac{1}{(n+1)!}\frac{1}{1-\left(\frac{1}{n+1}\right)}=\sum_{k=0}^{n}\frac{1}{k!}+\frac{1}{n!n}.$$

This tells us that $\sum_{k=0}^{n}\frac{1}{k!}<e<\sum_{k=1}^{n}\frac{1}{k!}+\frac{1}{n!n}$, so we can calculate how many terms we would need to calculate to get an error less than any particular bound. For instance, if we allow the series to go to 26 terms (*n*=25), we know that the total error is less than 2.5×10^{-27}.

As we've seen, this form of the definition of *e* has very nice convergence properties, so let's take the opportunity to calculate an approximation for *e*:

n	$\frac{1}{n!}$	$\sum_{k=0}^{n}\frac{1}{k!}$
0	1.000000000000000000000000	1.000000000000000000000000
1	1.000000000000000000000000	2.000000000000000000000000
2	0.5000000000000000000000000	2.500000000000000000000000
3	0.1666666666666666666666667	2.666666666666666666666667
4	0.0416666666666666666666667	2.708333333333333333333333
5	0.0083333333333333333333333	2.716666666666666666666667
6	0.0013888888888888888888889	2.718055555555555555555556
7	0.0001984126984126984126984	2.718253968253968253968254
8	0.0000248015873015873015873	2.718278769841269841269841
9	0.0000027557319223985890653	2.718281525573192239858907
10	0.0000002755731922398589065	2.718281801146384479717813
11	0.0000000250521083854417188	2.718281826198492865159532
12	0.0000000020876756987868099	2.718281828286168563946342
13	0.0000000001605904383682161	2.718281828446759002314558
14	0.0000000000114707455977297	2.718281828458229747912288
15	0.0000000000007647163731820	2.718281828458994464285470
16	0.0000000000000477947733239	2.718281828459042259058793
17	0.0000000000000028114572543	2.718281828459045070516048
18	0.0000000000000001561920697	2.718281828459045226708117
19	0.0000000000000000082206352	2.718281828459045234928753
20	0.0000000000000000004110318	2.718281828459045235339784
21	0.0000000000000000000195729	2.718281828459045235359357
22	0.0000000000000000000008897	2.718281828459045235360247
23	0.0000000000000000000000387	2.718281828459045235360286
24	0.0000000000000000000000016	2.718281828459045235360287
25	0.0000000000000000000000000	2.718281828459045235360287

Because the series 1e exists, this means that the limit 1c also exists, as we found the definition 1e was equivalent to definition 1c (even if they do look quite different).

Let's try to apply the same sort of procedure to definition 1d, and see if the same thing pops out:
We are still limited to the integers, but we can't use the same form of the binomial theorem because we're raising our expression to a negative exponent. As such, we'll need a more general form of the binomial theorem (often called the extended binomial theorem) that allows for any real exponent:
$(x+a)^r = \sum_{k=0}^{\infty}\binom{r}{k}x^k a^{r-k}$. We'll use a variation of this, $(1-x)^{-n} = \sum_{k=0}^{\infty}\binom{k+n-1}{k}x^k$ (which can be extracted using the identity $\binom{n}{k} = (-1)^k\binom{k-n-1}{k}$).

Our goal is now to justify a bit of Analytic Judo that we want to be able to do:
$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{-n} = \lim_{n\to\infty}\lim_{j\to\infty}\sum_{k=0}^{j}\binom{k+n-1}{k}\frac{1}{n^k} \overset{?}{=} \lim_{j\to\infty}\lim_{n\to\infty}\sum_{k=0}^{j}\binom{k+n-1}{k}\frac{1}{n^k}$. Are we allowed to exchange the limits in this particular case? We'll argue that we can through a somewhat technical Analytic argument.

First, we'll revisit definition 1d, but with a twist. In order to justify some later manipulation, we'll leave our formula as a power series, so that we can establish its convergence properties more easily:
$(1-x)^{-n} = \sum_{k=0}^{\infty}\binom{k+n-1}{k}x^k = \sum_{k=0}^{\infty}\frac{(k+n-1)^{\underline{k}}}{k!}x^k = \sum_{k=0}^{n}\underbrace{\left[\frac{(k+n-1)(k+n-2)...(n)}{k!}\right]}_{a_k}x^k$. This is not the most pleasant thing in the world, but recall that we can establish the radius of convergence, *r*, of this power series by looking at $r = \lim_{k\to\infty}\left|\frac{a_k}{a_{k+1}}\right|$.

In this case $r = \lim_{k\to\infty}\left|\frac{(k+n-1)(k+n-2)...(n)}{k!}\right|\left|\frac{(k+1)!}{(k+n)(k+n-1)...(n)}\right| = \lim_{k\to\infty}\frac{(k+1)}{(k+n)} = 1$, so we know that this function has a radius of convergence of 1 (about *x*=0). We also know that within any closed interval which is a subset of the interval $(-1,1)$, this function converges uniformly.

Now let's apply this information to our area of interest:
$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{-n} = \lim_{n\to\infty}\sum_{k=0}^{\infty}\binom{k+n-1}{k}\frac{1}{n^k}$. As long as $n>1$, we are in the interval where our power series converges uniformly.

We now want to establish that for each fixed *j*, the series converges (pointwise convergence is acceptable here). We have $\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{-n}=\lim_{n\to\infty}\lim_{j\to\infty}\underbrace{\sum_{k=0}^{j}\binom{k+n-1}{k}\frac{1}{n^k}}_{b_{n,j}}$.

Now, we can notice that for each fixed *j*, we have

$$\lim_{n\to\infty}\sum_{k=0}^{j}\left[\frac{(k+n-1)(k+n-2)\dots(n)}{k!}\right]\frac{1}{n^k}=\sum_{k=0}^{j}\lim_{n\to\infty}\left[\frac{(k+n-1)(k+n-2)\dots(n)}{n^k}\right]\frac{1}{k!}=\sum_{k=0}^{j}\frac{1}{k!}$$

(as a fixed *j* just leaves a standard polynomial in *n*).

The above shows us that the limit exists for each fixed *j*, and that we have uniform convergence when we fix $n>1$. This tells us that we can exchange our limits without affecting our summation. Let's do so:

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{-n}=\lim_{n\to\infty}\lim_{j\to\infty}\sum_{k=0}^{j}\binom{k+n-1}{k}\frac{1}{n^k}=\lim_{j\to\infty}\sum_{k=0}^{j}\lim_{n\to\infty}\left[\frac{(k+n-1)(k+n-2)\dots(n)}{n^k}\right]\frac{1}{k!}=\sum_{k=0}^{\infty}\frac{1}{k!}$$

This shows that the definition 1d converges to the same limit as definition 1c, which as stated previously means that the limit in definition 1b exists.

We have now justified all the definitions of the constant *e*, so let's return to our original definition which involved the function $f(x)=e^x$ and its unusual derivative.

This relationship can also be used to derive a power series expansion for e^x, using the Taylor Series expansion about 0 (also called the Maclaurin Series):

$$f(x)=\sum_{k=0}^{\infty}\frac{f^{(k)}(0)}{k!}x^k$$

In this case, with $f(x)=e^x$, we have the delightful case where $f^{(k)}(x)=e^x$ and so $f^{(n)}(0)=e^0=1$, leaving

Def 1f: $e^x=\sum_{k=0}^{\infty}\frac{x^k}{k!}$.

It is fairly easy to demonstrate that this power series possesses nearly all characteristics good: it is absolutely convergent, uniformly convergent and uniformly continuous for any bounded interval, and it has a cherry on top for all real (and for that matter, complex) values of *x*.

We can also use the extended binomial theorem to suggest another limit formula for e^x. If we proceed using the same reasoning as we used to connect definitions 1d and 1e, we would get a chain of equalities that looked like this:

$$\sum_{k=0}^{\infty}\frac{x^k}{k!}=\sum_{k=0}^{\infty}\lim_{n\to\infty}\left[\frac{n^k}{k!}\left(\frac{x}{n}\right)^k\right]\overset{?}{=}\lim_{n\to\infty}\sum_{k=0}^{\infty}\frac{n^k}{k!}\left(\frac{x}{n}\right)^k=\lim_{n\to\infty}\sum_{k=0}^{\infty}\binom{n}{k}\left(\frac{x}{n}\right)^k=\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n.$$

The only uncertain bit here (and it's a fairly significant uncertain bit!) is whether the "?" flagged exchange of limits is justified. We established this previously by looking at the convergence properties of the power series, doing some analysis theorem magic, and showing that we were allowed to exchange limits. Rather than subject ourselves to that again (with another variable to keep track of!), let's use another approach.

We'll attempt to demonstrate that the right hand side equals the left hand side:
Let $x_n = \left(1+\frac{x}{n}\right)^n = \sum_{k=0}^{n}\binom{n}{k}\left(\frac{x}{n}\right)^k$ and $s = \sum_{k=0}^{\infty}\frac{x^k}{k!}$. We want to demonstrate that these two things approach each other as n gets big, so we'll treat this as a limit: We want to show that $|s - x_n| \to 0$ as $n \to \infty$.

To state this in a different way: we want to show that that x_n and s are equal to each other. In order to do this, we will show that, no matter how close we need the two quantities to be, we can force the sequence to be that close (or closer) for all portions of the sequence after some particular position in the sequence. We'll do this by showing that no matter how close we want the sequence to be to the series (we'll call this distance ε), we can choose a large enough value N such that if $n \geq N$ then $|s - x_n| < \varepsilon$.

In order to demonstrate this, we'll look at two places where we could get "distance" between s and x_n: after some large value, m, and before this large value m (we'll establish what m is shortly).

Given a particular value for x (fix x), and selecting a particular value for $\varepsilon > 0$, we can select a particular m such that $\frac{|x|^{m+1}}{(m+1)!} + \frac{|x|^{m+2}}{(m+2)!} + \frac{|x|^{m+3}}{(m+3)!} + \ldots < \frac{\varepsilon}{2}$ (if this seem an unreasonable request, remember we can over-estimate this sum using a geometric sum:

$$\frac{|x|^{m+1}|x|^0}{(m+1)!} + \frac{|x|^{m+1}|x|^1}{(m+1)!(m+2)} + \frac{|x|^{m+1}|x|^2}{(m+1)!(m+2)(m+3)} + \ldots <$$

$$\frac{|x|^{m+1}|x|^0}{(m+1)!} + \frac{|x|^{m+1}}{(m+1)!}\left(\frac{|x|}{m+2}\right)^1 + \frac{|x|^{m+1}}{(m+1)!}\left(\frac{|x|}{m+2}\right)^2 + \ldots + \frac{|x|^{m+1}}{(m+1)!}\left(\frac{|x|}{m+2}\right)^k =$$

$$\frac{|x|^{m+1}(m+2)}{(m+1)!(m+2-|x|)} < \frac{|x|^{m+1}(m+2)}{(m+1)!}$$

We can make this less than $\frac{\varepsilon}{2}$ by selecting m to be appropriately large (with fixed x).

For some n>2, we have
$x_n = \left(1+\frac{x}{n}\right)^n = \sum_{k=0}^{n}\binom{n}{k}\left(\frac{x}{n}\right)^k = 1 + \binom{n}{1}\frac{x}{n} + \ldots + \binom{n}{k}\frac{x^k}{n^k} + \ldots + \binom{n}{n}\frac{x^n}{n^n}$. Examining an arbitrary term, we see that

$$\binom{n}{k}\frac{x^k}{n^k}=\frac{n^k x^k}{k!n^k}=\frac{x^k}{k!}\frac{(n)}{n}\frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-k+1)}{n}=\frac{x^k}{k!}\left[(1)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right],$$

so we can rewrite our series as

$$x_n=1+\frac{x}{1!}\left[(1)\right]+\frac{x^2}{2!}\left[(1)\left(1-\frac{1}{n}\right)\right]+\ldots+\frac{x^k}{k!}\left[(1)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]+\ldots$$

The *k*th term's coefficient is positive and smaller than $\frac{1}{k!}$ (as each term in $\left[(1)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]$ is positive and less than 1). Subtracting:

$$|s-x_n|=(1-1)+(1-1)+\left[\frac{|x|^2}{2!}-\frac{|x|^2}{2!}\left(1-\frac{1}{n}\right)\right]+\ldots+\left[\frac{|x|^k}{k!}-\frac{|x|^k}{k!}\left[(1)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]\right]+\ldots$$

For *n*>*m*, we see that we can stop at the *m*th term, and just correct using the error term that we had calculated:

$$|s-x_n|<\frac{|x|^2}{2!}\left[1-\left(1-\frac{1}{n}\right)\right]+\ldots+\frac{|x|^m}{m!}\left[1-\left[(1)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right)\right]\right]+\frac{\varepsilon}{2}.$$

Looking at the kth term (again, remember, we've fixed *x*), we see that for sufficiently large selection of n, we can cause each term to be arbitrary small, thus we can force the entire sum of *m* terms to be less than $\frac{\varepsilon}{2}$. As such, we have found that $|s-x_n|<\varepsilon$ for sufficiently large *n*, and so we have $\sum_{k=0}^{\infty}\frac{x^k}{k!}=\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$, bringing us to our next definition.

Def 1g: $e^x=\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$

Now that we have this (limit form) definition for e^x, we can substitute and get one for $e^{-x}=\frac{1}{e^x}=\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n$. We could then just raise both sides to the power -1, leaving us with our next definition.

Def 1h: $e^x=\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^{-n}$

All of these definitions were ultimately related to Definition 1, having to do with its own derivative.

1.2 The Real Logarithm

We start with the basic definition of the logarithm:

Def 2: $\log_a x = z \Leftrightarrow a^z = x$

So, the logarithm of *x* is the power we would need to raise *a* to, in order to get *x*.

From this definition, it is reasonably clear that $\log_a a^x = x$ (so, these two functions are compositional inverses by definition), which immediately brings us to the second definition of the logarithm

Def 2a: If $f\left(x\right)=a^x$ then $f^{-1}\left(x\right)=\log_a x$. Thus, $f^{-1}\left(f\left(x\right)\right)=f\left(f^{-1}\left(x\right)\right)=x$.
In other words, the logarithm can be defined as the compositional inverse of the exponential function.

As an aside, with this in mind, it is easy to prove the basic logarithm identities.

In each of these digressions, we restrict our values of a, b, c and *x* to real positive values.

Digression 1: $\log_a\left(xy\right)=\log_a\left(x\right)+\log_a\left(y\right)$
First, note that $f\left(x\right)=a^x$ is a one-to-one function in the reals, so $a^x = a^y \Leftrightarrow x = y$.

$$\log_a\left(xy\right)=\log_a\left(x\right)+\log_a\left(y\right)\Leftrightarrow a^{\log_a(xy)}=a^{\log_a(x)+\log_a(y)}=a^{\log_a(x)}a^{\log_a(y)}\Leftrightarrow xy=xy$$

Digression 2: $\log_a\left(x^b\right)=b\log_a\left(x\right)$
Again, note that $f\left(x\right)=a^x$ is a one-to-one function in the reals.

$$\log_a\left(x^b\right)=b\log_a\left(x\right)\Leftrightarrow a^{\log_a\left(x^b\right)}=a^{b\log_a(x)}=\left(a^{\log_a(x)}\right)^b\Leftrightarrow x^b=x^b$$

Digression 3: $\log_a\left(x\right)=\dfrac{\log_b\left(x\right)}{\log_b\left(a\right)}$
Looking at the left hand side, by our definition of the logarithm:
(1) $\log_a\left(x\right)=z\Leftrightarrow a^z=x$
Now note that we can write *a* in terms of *b* raised to some constant *c*, where *c* is a positive real constant:
(2) $a=b^c$
So, by (2) $a^z=\left(b^c\right)^z=b^{cz}=x$.

By our definition of the logarithm, $\log_b\left(x\right)=cz$, thus $\dfrac{\log_b\left(x\right)}{c}=z$.

Remembering that *(1)* tells us that $z=\log_a\left(x\right)$, we have $\dfrac{\log_b\left(x\right)}{c}=\log_a\left(x\right)$.

But what is *c*? From *(2)*, $a=b^c\Leftrightarrow\log_b\left(a\right)=c$, so $\log_a\left(x\right)=\dfrac{\log_b\left(x\right)}{\log_b\left(a\right)}$.

Digression 4: $\log_b(a) = \dfrac{1}{\log_a(b)}$

One last time, note that $f(x) = a^x$ is a one-to-one function in the real numbers.

$\log_b(a) = \dfrac{1}{\log_a(b)} \Leftrightarrow b^{\log_b(a)} = b^{\frac{1}{\log_a(b)}}$, or equivalently,

$a = b^{\frac{1}{\log_a(b)}} \Leftrightarrow a^{\log_a(b)} = b^{\frac{\log_a(b)}{\log_a(b)}} = b \Leftrightarrow b = b$, which is obviously true.

Now, given these definitions, we can explore the behavior of the logarithm somewhat. First, let's do a shortcut through some calculus landscape. We know that the exponential function is quite well behaved: it's continuous, one-to-one, and differentiable everywhere in its domain, and it never equals 0. This has a nice consequence for the compositional inverse, the logarithm. It means that the logarithm is also a differentiable function throughout its domain.

Let's look for the derivative of the logarithm function. Take $f(x) = a^x$ and $f^{-1}(x) = \log_a x$. We know that the logarithm and the exponentiation functions are compositional inverses, thus $f\left(f^{-1}(x)\right) = x$. Taking the derivative of both sides using the chain rule (which we are allowed to do, as we know that both the exponential and the logarithm functions are differentiable):

$\dfrac{d}{dx}\left[f\left(f^{-1}(x)\right)\right] = \dfrac{d}{dx}[x] = 1$, or $f'\left(f^{-1}(x)\right)\dfrac{d}{dx}\left[f^{-1}(x)\right] = 1$, thus $\dfrac{d}{dx}\left[f^{-1}(x)\right] = \dfrac{1}{f'\left(f^{-1}(x)\right)}$. In our case (because *f* is an exponential function), we know that $f'(x) = f(x)\,f'(0)$, so $\dfrac{d}{dx}\left[f^{-1}(x)\right] = \dfrac{1}{f'(0)\,f\left(f^{-1}(x)\right)} = \dfrac{1}{f'(0)\,x}$.

That's nice, in that it opens a very nice geometric meaning for us, involving the definite integral. First a bit of setup.

Let $L(x) = \int \frac{1}{x}dx$. It should be clear that this could be related to the logarithm of some base, but it isn't clear which one. The hint is that again, this is true in the case where $f'(0) = 1$, which was the same magic as when we were looking for *e* above, and so $L(x) = \log_e(x) + C$. And so, we get our next definition:

Def 2b: $\log(x) = \int_1^x \frac{1}{t}dt$

(Note, we are using $\log(x)$ to mean log base *e*. Some mathematicians consider using *ln* the moral equivalent of pronouncing the name "Leonhard Euler" such that it rhymes with "Ferris Bueller".)

This leads to (yet another) definition for *e*:

Def 1g: *e* is the unique number such that $1=\int_1^e \frac{1}{t}dt$

Back to arbitrary bases! As $\frac{d}{dx}\left[f^{-1}(x)\right]=\frac{1}{f'(0)x}$, we can use the fundamental theorem of calculus to connect our logarithm idea to the area under the curve of the function $\frac{1}{x}$: first, we're going to need to figure out what our log function is going to be at 1. Note that $a^0=1$ implies that $\log_a(1)=0$. Subtracting $\log_a(1)$ doesn't change our expression, so $\log_a x=\frac{1}{f'(0)}\int_1^x \frac{1}{t}dt=\frac{1}{f'(0)}\left[L(x)-L(1)\right]=\frac{\log(x)}{f'(0)}$.

This implies that all logarithms are related to the area under the function $\frac{1}{x}$, with only some (positive, real) scaling constant that depends only on the base of the logarithm applied.

We've been studiously averting our eyes from this constant for long enough. We will now rise up and determine what this constant is.

Recall that we started looking at exponential functions; that is, functions of the form $f(x)=a^x$, where *a* is some positive real number. We obtained some idea about derivatives of these functions in general: $f'(x)=f(x)f'(0)$. We then effectively defined *e* as the constant that caused $f'(0)=1$. So now we know how to take the derivative of exponentials base *e*, but other bases aren't that clear. Let's use the same trick that we used for Digression 3: we'll say that we can write any positive real *a* value as a power of *e*, $a=e^c$ (where *c* is a real, positive constant). Now note that this is the definition of the logarithm base *e*, so $c=\log(a)$. Now $f(x)=e^{x\log(a)}$. We can take the derivative of this now fairly easily, using the chain rule: $f'(x)=\log(a)e^{x\log(a)}=\log(a)f(x)$. We already knew that $f'(x)=f(x)f'(0)$, so that means that $f'(0)=\log(a)$.

And so, we finally get answers to our more general questions:

Digression 5: $\frac{d}{dx}\left[a^x\right]=\log(a)a^x$

Digression 6: $\frac{d}{dx}\left[\log_a\left(x\right)\right]=\frac{1}{\log\left(a\right)x}$

1.3 Applications of *the real exponential function (stub)*

1.3.1 Differential equations (stub)

1.3.1.1 The Catenary (stub)

1.3.1.2 Continuously Compounding Interest (stub)

Note moneychangers, compound interest formula.

1.3.1.3 Radioactive Decay (stub)

1.3.1.4 Capacitor and Inductor Charge and Discharge (stub)

1.3.1.5 Unbounded Population Growth Decay (stub)

chemical reaction rates
- rate of diffusion
  - pollutants
  - medication absorption / elimination
- 1st order kinetics
- binding of antigens to receptors
- muscle oxygen consumption
- hydration of cement

1.3.2 Statistics (stub)

normal
- Central Limit Theorem

Poisson
Gamma
chi-squared

1.3.3 unsorted (stub)

All Things Fourier
Laplaz transform
marriage proposals (game theory)

2 The Complex Numbers Defined

This paper so far has been (to quote Michael Spivak) "unremitting propaganda for the real numbers". Now is the time to remedy this inadequacy. *e* has a central role to play in the complex numbers, but its role can't be appreciated until we talk about what the complex numbers are.

Little in the standard up-to-calculus college mathematics curriculum provides as fruitful a source of confusion as the complex numbers. This is so enshrined into the culture of mathematics that even the names for this system suggest that something is amiss: "imaginary" numbers have no "real" component, for instance. (So, these components are then "fake"?) Truthfully, both the real numbers and the imaginary numbers are in some sense artificial. Despite this, the complex numbers are normally sequestered away in the mathematics curriculum until they stumble onto the stage during an elementary algebra class. The standard introduction to this concept generally proceeds as follows:

1. $i = \sqrt{-1}$. Yes, this was something that you couldn't previously do, but now we're going to. Don't like it? Yes, well, all life is suffering (or perhaps "life is filled with a sense of pervasive unsatisfactoriness"?), get over it. Oh, by the way, there's a test in a week.
2. (Wave hands)
3. Look how convenient this is! (Profit!)

There are a few reasons that the subject is dealt with this way. The primary reason is that it's time intensive to actually explain what is going on, why it evolved historically, or even some of the underlying structure that enables the profoundly powerful techniques that rest on the complex numbers. The formalism (read: justification) underlying the complex numbers is fairly modern mathematics (early 19th century, which post-dates the bulk of math that most people are ever exposed to). In addition, the original development of the complex numbers was remarkably similar to how current-day students are introduced: it was informal, not terribly well understood, and principally motivated by expediency.

What follows is a penny tour of the ideas that underlie the complex number. We'll return to the significance of *e* in the complex numbers shortly.

The motivation for the complex numbers is straight forward enough. We know that certain quadratic equations don't have real roots. We know this because every quadratic polynomial is actually a parabola in disguise (and that every parabola is a quadratic polynomial). Of course, many parabolas never intersect the x-axis, so it follows that not every quadratic polynomial should have roots.

Algebraically, we found that we could get nice, closed form solutions for the roots of quadratic polynomials through the quadratic equation:
Equations of the form $0 = ax^2 + bx + c$, $a \neq 0$ have a very nice closed form solution that we can derive easily.

We divide both sides of our equation by *a*: $0 = x^2 + \frac{b}{a}x + \frac{c}{a}$. Rearranging, we get $-\frac{c}{a} = x^2 + \frac{b}{a}x$.

Now, the magic step: we complete the square on the right hand side by adding $\frac{b^2}{4a^2}$ to both sides, leaving us with $\frac{b^2}{4a^2} - \frac{c}{a} = x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}$.

The whole point to completing the square was so we could write the right hand side as a perfect square, but we can also simplify the left hand side using a common denominator: $\frac{b^2 - 4ac}{4a^2} = \left(x + \frac{b}{2a}\right)^2$.

Now, we take the square root of both sides: $\pm\sqrt{\frac{b^2 - 4ac}{4a^2}} = x + \frac{b}{2a}$. The denominator of the left hand side can be simplified, and we can move the $\frac{b}{2a}$ term to the left hand side: $\pm\frac{\sqrt{b^2 - 4ac}}{2a} - \frac{b}{2a} = x$.

Simplifying the right hand side, we get the standard presentation for the quadratic equations: $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

In this form, the discriminant ($b^2 - 4ac$) establishes how many solutions there are. When the discriminant is less than zero, we don't get an answer that we can deal with, because (without *i*) we don't know how to take the square root of negative numbers.

It would be nice to be able to solve this sort of equation. Why? The actual answer requires that we get a bit further, so we'll come back to this question. For now, we'll content ourselves with the idea that a general solution would give us the ability to solve those pesky quadratics that don't intersect the x-axis and deal with the (not-quite-as-famous-as-the-quadratic-formula) cubic formula, which can have complex intermediate values, even in the case where real roots do exist. This last reason was Bombelli's initial inspiration in his study of complex numbers, which was the first serious attempt to understand the complex numbers (in the 16th century).

So, we want to develop a number system that looks as much like the real numbers as possible, but then has some extra stuff that allows us to find roots in all the quadratics.

2.1 Fields and Field Extensions

Let's try to get some idea of the important characteristics we want to retain. The real numbers form a field (which is unrelated to what physicists call a field). This means that most of the rules that we have come to know and love associated with the real numbers apply. A field, *K* (*a*, *b* and *c* are arbitrary field elements):

- is closed under both addition and multiplication ($(a+b)\in K$, $(ab)\in K$)
- is associative in addition and multiplication (i.e., $a(bc)=(ab)c$ and $a+(b+c)=(a+b)+c$)
- is commutative in addition and multiplication (i.e., $ab=ba$ and $a+b=b+a$)
- has an identity element for addition, called *0*, and an identity element for multiplication, called *1* (i.e., $1a=a$ and $0+a=a$)
- has additive inverses for all elements and multiplicative inverses for all elements other than 0 (i.e. $a+(-a)=0$ and $aa^{-1}=1$)
- is distributive (i.e., $a(b+c)=ab+ac$

A ring is similar to a field, except that the multiplication operator need not be commutative, there need not be a multiplicative identity, and (even in the case where a multiplicative identity exists) multiplicative inverses don't necessarily exist for some non-zero elements. In some cases, we will also be interested in a refinement of this idea called a commutative ring with unity (CRU) , which is largely equivalent to a field, except that some (non-zero) elements in the set may not have multiplicative inverses. In a CRU, non-zero elements that have multiplicative inverses are called units.

So we want to add something to the real numbers, and keep all those characteristics. Well, we can't just go adding a single value, as we need to be concerned what will happen when we multiply or add things to the additional element (Remember, a field is closed!) As a matter of notation, we'll call this process "extending a field". When we extend a field, we are making a new field that contains our original field as a subset, using operators that act the same way in the original field. In this case, we're looking for the minimal extension field that contains the original field along with the element that we're adding (this field is minimal in the sense that every field that contains both our original field and this element contain our minimal field). If we're adding the element α to the field *K*, we'll call this "adjoining α to *K*", and write this minimal field $K(\alpha)$.

2.2 An example extension: $\mathbb{Q}(\sqrt{2})$

Let's suppose that you were the first person in ancient Greece to discover that $\sqrt{2}$ wasn't a rational number. After demonstrating your lovely proof, you are surrounded by angry mathematicians who don't think that irrational numbers are meaningful, and who are incensed that you would suggest that a shape as simple as a right triangle might not exist. In the 10 minutes that you have prior to being drowned by your peers, you need to attempt to mollify them by adding the element $\sqrt{2}$ to the rational numbers, while retaining the field properties of the rational numbers (thus making $\mathbb{Q}\left(\sqrt{2}\right)$).

Well, as a first guess, we know that we need the rational numbers, and we know that we'll need all the rational numbers multiplied by $\sqrt{2}$: for our discussion, let's name the proposed minimal extension field $A=\left\{a+b\sqrt{2}:a,b\in\mathbb{Q}\right\}$. Well, this is certainly a superset of the rational numbers (if *b*=0, then we get the rational numbers). Is $\sqrt{2}$ in the set? Certainly! If *a*=0 and *b*=1, we have $\sqrt{2}$. Now we need to verify that this forms a field (under the addition and multiplication operators borrowed from the real numbers) and that this is the minimal field that contains both the rational numbers ($\mathbb{Q}$) and $\sqrt{2}$.

A Forms a Field

We get most of the field properties for *A* (everything but closure and inverses) through the field properties of $\mathbb{R}$. We can demonstrate *A* is closed under addition and multiplication by just doing the algebra on two elements in *A*; we'll call the two elements $a+b\sqrt{2}$ and $c+d\sqrt{2}$, where *a*, *b*, *c*, and *d* are rational numbers.

$\left(a+b\sqrt{2}\right)\left(c+d\sqrt{2}\right)=\left(ac+2bd\right)+\left(ad+bc\right)\sqrt{2}$ and

$\left(a+b\sqrt{2}\right)+\left(c+d\sqrt{2}\right)=\left(a+c\right)+\left(b+d\right)\sqrt{2}$.

$ac+2bd$, $ad+bc$, $a+c$, and $b+d$ are all rational numbers, so we do have closure. Additive inverses are easy (just multiply by -1). Multiplicative inverses are also fairly straight forward given a bit of algebra: if we want to find the multiplicative inverse for $\left(a+b\sqrt{2}\right)$, then we are looking for $\frac{1}{\left(a+b\sqrt{2}\right)}$. Rationalizing the denominator, we get $\frac{1}{\left(a+b\sqrt{2}\right)}1=\frac{1}{\left(a+b\sqrt{2}\right)}\frac{\left(a-b\sqrt{2}\right)}{\left(a-b\sqrt{2}\right)}=\frac{\left(a-b\sqrt{2}\right)}{\left(a^2+2b^2\right)}$. Written in a different form $\left(\frac{a}{\left(a^2+2b^2\right)}+\frac{-b}{\left(a^2+2b^2\right)}\sqrt{2}\right)$, it is clear that the multiplicative inverse is also in *A*, so there's our answer!

This prior paragraph demonstrated that *A* is a field that contains both $\mathbb{Q}$ and $\sqrt{2}$. We know that it must contain $\mathbb{Q}\left(\sqrt{2}\right)$ as a subfield, by the definition of $\mathbb{Q}\left(\sqrt{2}\right)$. Consequently, we know that $\mathbb{Q}\left(\sqrt{2}\right)\subseteq A$.

A is Minimal

In order to argue that it is, we are going to use a proof by contradiction; we'll assume that it isn't minimal (that is, we'll assume that $\mathbb{Q}\left(\sqrt{2}\right)$ is a proper subset of *A*, and thus that there is some element in *A* that is not in $\mathbb{Q}\left(\sqrt{2}\right)$). We will then show that under this assumption we arrive at a demonstratively false statement (the contradiction). This contradiction indicates that our assumption that *A* was not minimal was incorrect, which means that *A* is minimal.

We assume that *A* is not minimal, thus there is some "extra" member in *A* that doesn't need to be in $\mathbb{Q}\left(\sqrt{2}\right)$; we'll call this extra member $a+b\sqrt{2}$, where *a* and *b* are rational numbers (we know that we can write the element in this way, because all elements in *A* can be written in this way). We know that the actual minimal extension field has to include $\sqrt{2}$ (by our definition of $\mathbb{Q}\left(\sqrt{2}\right)$), so we'll start with this element, and we'll multiply it by *b* (which is a rational number, and thus also in $\mathbb{Q}\left(\sqrt{2}\right)$). By the closure property of the field $\mathbb{Q}\left(\sqrt{2}\right)$, the result needs to be $\mathbb{Q}\left(\sqrt{2}\right)$ (so $b\sqrt{2}\in\mathbb{Q}\left(\sqrt{2}\right)$). Now let's add the rational number *a*. Again, this needs to be in $\mathbb{Q}\left(\sqrt{2}\right)$ (by the closure property), so that means that $a+b\sqrt{2}\in\mathbb{Q}\left(\sqrt{2}\right)$. This is a contradiction, as we had assumed that $a+b\sqrt{2}\notin\mathbb{Q}\left(\sqrt{2}\right)$.

Unhappy Endings

Now that we have demonstrated that *A* is a field, and that *A* is minimal, we know that $\mathbb{Q}\left(\sqrt{2}\right)=\left\{a+b\sqrt{2}:a,b\in\mathbb{Q}\right\}$. Sadly, the enraged Greek mathematicians still drown you.

Despite the historical drowning, this is still an exciting find! Sadly, it doesn't always work out this nicely. $\mathbb{Q}\left(\sqrt[3]{2}\right)$ isn't so nice, for instance; were it the same construction, we would expect $\mathbb{Q}\left(\sqrt[3]{2}\right)\overset{?}{=}\left\{a+b\sqrt[3]{2}:a,b\in\mathbb{Q}\right\}$. In this case, closure for multiplication runs into a problem: $\left(a+b\sqrt[3]{2}\right)\left(c+d\sqrt[3]{2}\right)=\left(ac\right)+\left(ad+bc\right)\sqrt[3]{2}+bd\sqrt[3]{4}$. That's great, but we don't have any way of making the $\sqrt[3]{4}$ term from rational numbers, or a rational number multiplied by $\sqrt[3]{2}$. So, we find that our proposed set isn't a field at all (as it is not closed under multiplication). As such, the complete answer has to be more interesting than our initial attempt would indicate.

Prior to moving on, it is useful to point out that we've been extending the rational numbers here. $\mathbb{R}\left(\sqrt[3]{2}\right)=\mathbb{R}$, because $\sqrt[3]{2}\in\mathbb{R}$, thus the minimal field containing both $\sqrt[3]{2}$ and $\mathbb{R}$ is $\mathbb{R}$! Extending the rational numbers makes sense, because it's clear what we could adjoin to $\mathbb{Q}$ (any irrational number!). It is less clear what we could adjoin to the real numbers; axiomatically we define the real numbers to include just about everything that we normally consider a number (that is, the real numbers are "complete").

We'll have to take a step back and hope that some additional theory can save us.

2.3 Kronecker's Theorem

Kronecker's Theorem to the rescue!

Kronecker's Theorem states that for any field *K* and non-constant polynomial over *K* (written $f(x) \in K[x]$, or $f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_2 x^2 + a_1 x + a_0$, where each coefficient a_j is a member of the field *K*), there is an extension of *K* that contains a root for $f(x)$. In other words, even if $f(x)$ doesn't have a root in *K*, we can make a field extension for *K* that does contain a root for $f(x)$. Showing how this works is going to require some significant development. We'll first go through a series of definitions that we'll need for the development of Kronecker's Theorem, and then we'll take a look at a fairly general proof.

Preparations
The Division Algorithm
A rather strange thing will play a central role in our quest: something mathematicians like to call "The Division Algorithm", but which everyone else calls "polynomial long division". The Division Algorithm just says that if we have two polynomials, *f* and *g*, we can divide *f* by *g* (as long as *g* isn't zero!), and get a quotient, *q*, and some remainder, *r*. Either $r = 0$, in the case where *g* evenly divides *f*, or $r \neq 0$ and the degree of *r* is less than the degree of *g*. In all cases, $f = qg + r$. A trusting reader may be willing to accept this because of their prior experience with polynomial long division, but I'll provide a proof anyway.

If $g \mid f$ (*g* divides *f* evenly), then set $r = 0$, and *g* such that $f = qg$. If $g \nmid f$ (*g* does not divide *f* evenly), examine the set $R = \left\{ f - l\,\mathrm{g} : l \in K[x] \right\}$. We know that $0 \notin R$ (if it were, then *g* would have evenly divided *f*). Let *r* be the polynomial of least degree in *R*, and let *q* be the selection of *l* that produces that minimal polynomial. We assert that this fulfills our definitions. $f = qg + r$ by construction, but the fact that the degree of *r* is less than the degree of *g* isn't obvious. Let's proceed through a proof by contradiction: we'll assume that *r* is indeed the minimal degree element in *R*, and that the degree of *r* is greater than the degree of *g*. Let's assume that the highest degree term of *r* is ax^d and the highest degree term of *g* is bx^m. We have assumed that $d > m$. If we examine $\tilde{r} = r - ab^{-1}x^{d-m}g$, we note that we would subtract away the highest order term from r (so $\deg \tilde{r} < \deg r$). Also note that as $r = f - q\mathrm{g}$, $\tilde{r} = f - q\mathrm{g} - ab^{-1}x^{d-m}g = f - \left(q + ab^{-1}x^{d-m}\right)g$, which is of the proper form to be included in *R*. We had selected *r* to be of minimal degree in R, and $\deg \tilde{r} < \deg r$, so this is a contradiction. Thus, we know that the division algorithm always results in $r = 0$ or $r \neq 0$ and $\deg r < \deg g$.

Ideals
Now that we believe in polynomial division, let's define some notation. For $f(x) \in K[x]$, $\langle f(x) \rangle$ is called the ideal generated by $f(x)$ in $K[x]$, and it is defined as $\langle f(x) \rangle = \{g(x) f(x) : g(x) \in K[x]\}$. In other words, $\langle f(x) \rangle$ is the set of every polynomial that is evenly divisible by $f(x)$.

Equivalence Relations and Equivalence Classes
The next bit of preliminary information is quintessential mathematics: equivalence relations. Equivalence relations allow us to represent something like weak equality: equivalence relations are generally not literally equal (though, literal equality is actually an equivalence relation), but related in some weaker sense that we define (and which is hopefully in some way important to the matter at hand). A relation *R* on a set *S* is called an equivalence relation if and only if:

- $(a,a) \in R$ for all $a \in S$ (*R* is reflexive)
- $(a,b) \in R$ implies $(b,a) \in R$ for all $a,b \in S$ (*R* is symmetric)
- $(a,b) \in R$ and $(b,c) \in R$ implies $(a,c) \in R$ for all $a,b,c \in S$ (*R* is transitive)

A bit of thought on this concept establishes that an equivalence relation partitions its domain into some number of disjoint sets that completely cover that domain. The trivial (and uninteresting) cases are the identity relation (which results in every element being in its own partition) and the relation containing every possible pairing of elements of the set (which results in every element being in the same partition).

If we want to refer to the particular subset of the partition that contains an element *a*, we'll call this the equivalence class of *a*. We will denote two elements being in the same equivalence class as $a \sim b$ (where $a,b \in S$).

The standard equivalence relations that are common are the integers, modulo a positive integer. If we look at the integers, modulo 7, then there are exactly 6 disjoint equivalence classes, one corresponding to each of the integers 0 through 6. Each of these equivalence classes contain an infinite number of integers, each of which has the same remainder when divided by 7 (e.g., $3 \bmod 7 \sim 10 \bmod 7$).

The other common example of an equivalence relation that we use every day is the rational numbers. Each rational number has an infinite number of other rational numbers which are in the same equivalence class (e.g., $\frac{1}{2} \sim \frac{4}{8}$), and we think of each equivalence class as having one particular value.

Quotient Rings

The next notational issue is defining something called a quotient ring, $K\left[x\right]\big/\left\langle f\left(x\right)\right\rangle$.

This is defined as $K\left[x\right]\big/\left\langle f\left(x\right)\right\rangle = \left\{g\left(x\right)+\left\langle f\left(x\right)\right\rangle : g\left(x\right)\in K\left[x\right]\right\}$; this notation is reminiscent of division, so it shouldn't be too surprising that it plays a role. We'll setup an equivalence relation based on division. In this case, we'll define the equivalence class by saying that all elements that share the same remainder after being divided by $f\left(x\right)$ are "the same". For example, in $\mathbb{Q}\left[x\right]\big/\left\langle x^2-2\right\rangle = \left\{g\left(x\right)+\left\langle x^2-2\right\rangle : g\left(x\right)\in\mathbb{Q}\left[x\right]\right\}$, $0+\left\langle x^2-2\right\rangle \sim x^2-2+\left\langle x^2-2\right\rangle$ because both 0 and x^2-2 have a remainder of 0 when divided by x^2-2. Similarly, $1+\left\langle x^2-2\right\rangle \sim x^2-1+\left\langle x^2-2\right\rangle \sim x^4-3x^2+2+\left\langle x^2-2\right\rangle$.

The term "ring" in "quotient ring" implies that we can perform operations using elements of our quotient ring. We'll define the operations as follows:
$\left(g\left(x\right)+\left\langle f\left(x\right)\right\rangle\right)\times\left(h\left(x\right)+\left\langle f\left(x\right)\right\rangle\right)=g\left(x\right)h\left(x\right)+\left\langle f\left(x\right)\right\rangle$ and
$\left(g\left(x\right)+\left\langle f\left(x\right)\right\rangle\right)+\left(h\left(x\right)+\left\langle f\left(x\right)\right\rangle\right)=g\left(x\right)+h\left(x\right)+\left\langle f\left(x\right)\right\rangle$.

Irreducible Polynomials

And we're nearly there, but there's one more idea that we need to hit prior to addressing Kronecker's Theorem: irreducible polynomials. A polynomial of degree n, where n is greater than or equal to 1 (i.e., the largest power of *x* in the polynomial is 1 or greater) is called irreducible over a field *K* if we can not evenly divide out any other polynomial (over K) of degree 1 to *n*-1. For instance, for the polynomials over the real numbers x^2+1 is irreducible, because there aren't any linear terms that we can divide out (if there were real linear terms that we could divide out, then the polynomial would have a real root, but we know that it doesn't by the quadratic equation).

The Proof Begins

So, now we can finally proceed to Kronecker's Theorem. Given all this lead up, it seems reasonable to wonder what sort of rescue this is going to be when all is said and done...

To prove Kronecker's Theorem, we construct an extension field for *K* that contains a root for the polynomial $f\left(x\right)\in K\left[x\right]$. We'll create a construction for the proposed extension field, we'll demonstrate that it does form a field, we'll show that it is an extension field, and finally we'll show that we have an element in our extension field that is a root of our polynomial.

Proposed Construction

We'll approach this using a few cases:
case 1: $f\left(x\right)$ already has a root in K. For this case, we are already done! *K* can be used as its own "extension field" in this case.

case 2: $f(x)$ does not have a root in K, but is not irreducible over K. For this case, we'll select an irreducible factor of $f(x)$, and then use case 3 to make an extension field that contains a root of our irreducible factor. Because the new field contains a root of the selected irreducible factor, it contains a root for $f(x)$.

case 3: $f(x)$ *is irreducible over K, and does not have a root in K.* We claim that $K[x]/\langle f(x)\rangle$ is an extension field for *K* that contains a root for $f(x)$.

The Field Properties

To show that $K[x]/\langle f(x)\rangle$ is a field, we need to show that all the field properties hold. Most of the field properties are inherited from standard polynomial arithmetic: we get closure, associativity, commutativity, distributivity and the identities "for free" from the standard rules for polynomial arithmetic. We are left with the inverses. The additive inverse can be had by just multiplying the polynomial by -1.

The multiplicative inverse is the real problem here. We'll show that there is a multiplicative inverse for an arbitrary non-zero element, but we sadly won't really know what the element is, just that it exists.

We'll first take an arbitrary non-zero element from $K[x]/\langle f(x)\rangle$, which we'll call $a(x)+\langle f(x)\rangle$. We'll look at the cases where $a(x)+\langle f(x)\rangle$ is a constant polynomial, and where $a(x)+\langle f(x)\rangle$ is not a constant polynomial.

Case 1: $a(x)+\langle f(x)\rangle$ is a constant (or in the same equivalence class as a constant). We know that the constant can't be 0 (and that $f(x)$ can't divide $a(x)$); if it were, then our element would be in the same equivalence class as 0, which is inconsistent with how we chose $a(x)+\langle f(x)\rangle$. In this case, $a(x)+\langle f(x)\rangle$ has a multiplicative inverse in *K* (as *K* is a field) which will also work in $K[x]/\langle f(x)\rangle$.

Case 2: $a(x)+\langle f(x)\rangle$ is not a constant, and is not in the same equivalence class as a constant (note, we will effectively make the same argument as $\gcd(a(x), f(x))=1$).

We know that $f(x)$ is irreducible over *K*. We also know that if we multiply $a(x)+\langle f(x)\rangle$ by another element $b(x)+\langle f(x)\rangle$, the result looks like this:

$\left(a(x)+\langle f(x)\rangle\right)\left(b(x)+\langle f(x)\rangle\right)=a(x)b(x)+\langle f(x)\rangle=\underbrace{a(x)b(x)+g(x)f(x)}_{c(x)}$, where $g(x)\in K[x]$. We'll call this last expression $c(x)$.

Examine the set of possible values for $c(x)$, $C=\{a(x)b(x)+g(x)f(x):b(x),g(x)\in K[x]\}$. There are a set of non-zero polynomials of minimal degree in *C* (by the well ordered property of the positive integers). We'll select one of these minimal degree elements, which we'll call $\hat{c}(x)$. Because of the way that *C* is defined, we know that for some values of $\hat{b}(x)$ and $\hat{g}(x)$, $\hat{c}(x)=a(x)\hat{b}(x)+\hat{g}(x)f(x)$.

We will now argue that the ideal $\hat{C}=\langle\hat{c}(x)\rangle=\{\hat{c}(x)h(x):h(x)\in K[x]\}$ is equal to *C*. One containment is easy to see:
$d(x)\in\hat{C}$ implies that for some $\hat{h}(x)\in K[x]$, $d(x)=\hat{c}(x)\hat{h}(x)$.
$\hat{c}(x)=a(x)\hat{b}(x)+\hat{g}(x)f(x)$, so
$d(x)=\left(a(x)\hat{b}(x)+\hat{g}(x)f(x)\right)\hat{h}(x)=a(x)\hat{h}(x)\hat{b}(x)+\hat{g}(x)\hat{h}(x)f(x)$. We know that $\hat{h}(x)\hat{b}(x)\in K[x]$ and $\hat{g}(x)\hat{h}(x)\in K[x]$, so $d(x)\in C$, and thus $\hat{C}\subseteq C$.

The reverse containment is more problematic. If we let $d(x)\in C$, we know that for some value of $\tilde{b}(x),\tilde{g}(x)\in K[x]$ we can write $d(x)=a(x)\tilde{b}(x)+\tilde{g}(x)f(x)$. We also know by the division algorithm (standard polynomial division) we can also write $d(x)=q(x)\hat{c}(x)+r(x)$, where $q(x),r(x)\in K[x]$ and $\deg r(x)<\deg\hat{c}(x)$. Finally, we know that both $d(x)$ and $\hat{c}(x)$ are contained within *C*, so:

$$r(x)=d(x)-q(x)\hat{c}(x)=d(x)-q(x)\left(a(x)\hat{b}(x)+\hat{g}(x)f(x)\right)$$
$$=a(x)\tilde{b}(x)+\tilde{g}(x)f(x)-q(x)a(x)\hat{b}(x)-q(x)\hat{g}(x)f(x)$$
$$=a(x)\left(\tilde{b}(x)-q(x)\hat{b}(x)\right)+f(x)\left(\tilde{g}(x)-q(x)\hat{g}(x)\right)\in C.$$

This shows that $r(x)\in C$. We had selected $\hat{c}(x)$ to be of minimal degree in C, and $r(x)\in C$ and $\deg r(x)<\deg\hat{c}(x)$, so it must be the case that $r(x)=0$. It follows that $d(x)=q(x)\hat{c}(x)$, which is clearly in $\hat{C}$, so $C\subseteq\hat{C}$. These two containments imply that $C=\hat{C}$.

It isn't clear how that helps. We know that every member of C is also a member of $\hat{C}$, so let's look into the consequences. First, $1a(x)+0f(x)\in C=\hat{C}$, which tells us that for some $k(x) \in K[x]$, $a(x) = k(x)\hat{c}(x)$. Similarly, $0a(x) + 1f(x) \in C = \hat{C}$, which tells us that for some $h(x) \in K[x]$, $f(x) = h(x)\hat{c}(x)$, which implies that $\hat{c}(x)$ evenly divides $f(x)$. This certainly seems problematic, so let's look more closely.

We know that $f(x)$ is irreducible, so if $\hat{c}(x)$ divides $f(x)$, we know that $\hat{c}(x) = l \in K$ or $\hat{c}(x) = lf(x)$, $l \in K$.

Let's rule out one of these apparent choices: if $\hat{c}(x)$ is not a constant (and thus has degree ≥ 1) we run into a contradiction quite rapidly. We know that $a(x) = k(x)\hat{c}(x)$ and $\hat{c}(x) = lf(x)$, thus $a(x) = k(x)lf(x)$, which implies that $a(x) + \langle f(x) \rangle \sim 0 + \langle f(x) \rangle$, which is a contradiction (as we had selected $a(x) + \langle f(x) \rangle$ to be non-zero).

Thus, $\hat{c}(x)$ is a constant, which we'll call $\hat{c}(x) = l \in K$. We know that $\hat{c}(x) = l = a(x)\hat{b}(x) + \hat{g}(x)f(x)$, and we also know that $l \neq 0$ (as $a(x) + \langle f(x) \rangle$ is non-zero). This implies that we can multiply through the entire equation by l^{-1}, $l^{-1}\hat{c}(x) = l^{-1}l = a(x)l^{-1}\hat{b}(x) + l^{-1}\hat{g}(x)f(x) = 1$.

Hey, wait a minute! That's what we're looking for! This implies that $l^{-1}\hat{b}(x) + \langle f(x) \rangle$ is the multiplicative inverse for $a(x) + \langle f(x) \rangle$, (which we'll call $a(x)^{-1} + \langle f(x) \rangle$), and shows that our construction is a field.

The Construction is an Extension Field

Showing that $K[x]/\langle f(x) \rangle$ is an extension field of *K* is easy. Note that choosing a constant value as the polynomial yields that same constant as a remainder after division by $f(x)$. Because of the way that we defined addition and multiplication, we automatically inherit the field properties from *K*, so there is a subfield of $K[x]/\langle f(x) \rangle$ (the constant polynomials) which have the same behavior as *K*. Thus *K* is a subfield of $K[x]/\langle f(x) \rangle$.

$f(x)$ has a root in our Extension Field

Finally, showing that $f(x)$ has a root is easy from a formal setting. First note that the elements of $K[x]/\langle f(x) \rangle$ look like polynomials. We'll choose the element

$\alpha = x + \langle f(x) \rangle$ as our prospective root. $f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_2 x^2 + a_1 x + a_0$, so $f(\alpha) = a_n \left(x + \langle f(x) \rangle\right)^n + ... + a_1\left(x + \langle f(x) \rangle\right) + a_0$. The rules that we defined for arithmetic in our quotient ring allow us to simplify this greatly. Applying the multiplication rule, we get $f(\alpha) = \left(a_n x^n + \langle f(x) \rangle\right) + ... + \left(a_1 x + \langle f(x) \rangle\right) + a_0$. Finally, applying the addition rule, we get $f(\alpha) = a_n x^n + ... + a_1 x + a_0 + \langle f(x) \rangle = f(x) + \langle f(x) \rangle \sim 0 + \langle f(x) \rangle$, so in our quotient field, we have just found a root!

2.4 Adjoining Algebraic Numbers Through Kronecker's Theorem

The Minimal Polynomial

In order to draw a connection between Kronecker's Theorem and the style of field extensions that we were looking at earlier, we need to look into a small variation of irreducibility called the minimal polynomial.

The minimal polynomial for an algebraic number, α, over a field, *K*, is the monic polynomial (i.e., the polynomial with a leading coefficient of 1) over *K* which has α as a root, and which is of minimal degree. This is written $m_{\alpha,K}(x)$.

Sadly, we need to make this definition somewhat more formal for our task. Let's define a new set and show that it's an ideal. We'll let *I* be the set of all polynomials over K that have a root at the algebraic number α: $I = \{ f \in K[x] : f(\alpha) = 0 \}$ (because we're assuming that α is algebraic, we know that *I* is not empty). In this set of polynomials there is a subset of polynomials which have minimal degree (by the well ordered property of the integers). Select a monic polynomial of minimal degree. By the above definition, this polynomial is $m_{\alpha,K}(x)$. We assert that $I = \langle m_{\alpha,K}(x) \rangle$.

$\langle m_{\alpha,K}(x) \rangle = \{ m_{\alpha,K}(x) f(x) : f(x) \in K[x] \} \subseteq I$ fairly clearly: if an element $g(x) \in \langle m_{\alpha,K}(x) \rangle$, it is of the form $g(x) = m_{\alpha,K}(x) a(x)$ for some $a(x) \in K[x]$. $g(\alpha) = m_{\alpha,K}(\alpha) a(\alpha) = 0 a(\alpha) = 0$, so $g(x) \in I$, and $\langle m_{\alpha,K}(x) \rangle \subseteq I$.

The opposite inclusion requires another division algorithm argument: $g(x) \in I$. By the division algorithm, we get $g(x) = q(x) m_{\alpha,K}(x) + r(x)$, where $\deg r(x) < \deg m_{\alpha,K}(x)$. $g(x) \in I$ by assumption, $q(x) m_{\alpha,K}(x) \in I$ by construction, so that means that $g(x) - q(x) m_{\alpha,K}(x) \in I$, and thus $r(x) \in I$. As $m_{\alpha,K}(x)$ was chosen to have minimal degree, $r(x)$ could only equal 0, so we have $g(x) = q(x) m_{\alpha,K}(x)$, which means that $g(x) \in \langle m_{\alpha,K}(x) \rangle$ and thus $I \subseteq \langle m_{\alpha,K}(x) \rangle$.

To recap, we have found that $I=\langle m_{\alpha,K}(x)\rangle$. This has some amazing consequences for the minimal polynomial. The important ones for our purpose are:

1. The minimal polynomial is irreducible over *K*. Were it not, then $m_{\alpha,K}(x)=f(x)g(x)$, where both $\deg f(\alpha)\geq 1$ and $\deg g(\alpha)\geq 1$, and $m_{\alpha,K}(\alpha)=f(\alpha)g(\alpha)$, so either $f(\alpha)=0$ or $g(\alpha)=0$. $\deg m_{\alpha,K}(x)>\deg f(x)$ and $\deg m_{\alpha,K}(x)>\deg g(x)$, so this contradicts our choice of $m_{\alpha,K}(x)$ as the minimal polynomial.
2. If $g(x)\in K[x]$ and $g(\alpha)=0$ then $m_{\alpha,K}(x)$ divides $g(x)\in K[x]$. This is a direct consequence of $I=\langle m_{\alpha,K}(x)\rangle$.
3. If $g(x)\in K[x]$, $g(x)$ is monic and irreducible, and $g(\alpha)=0$ then $m_{\alpha,K}(x)=g(x)$. Note that by the previous finding, we knew that $m_{\alpha,K}(x)$ divides $g(x)$, but $g(x)$ is irreducible and monic, so they must be equal.

This implies that any monic, irreducible polynomial which has α as a root works! For example, $m_{\sqrt{2},\mathbb{Q}}(x)=x^2-2$, $m_{i,\mathbb{R}}(x)=x^2+1$ and $m_{\pi,\mathbb{R}}(x)=x-\pi$.

The interesting thing about this is that the minimal polynomial is unique (that is, for a given α, over a given field, there is only one possible $m_{\alpha,K}(x)$), but it is not exclusively the minimal polynomial for this element. For example, $m_{i,\mathbb{R}}(x)=x^2+1$, but it is also the case that $m_{-i,\mathbb{R}}(x)=x^2+1$. Elements that share the same minimal polynomial over a particular field are said to be conjugate over that field.

Adjoining Algebraic Numbers to Fields

Minimal polynomials provide the critical link between adjoining algebraic elements and Kronecker's Theorem.

First, let's think about adjoining an element to a field again. We had defined $K(\alpha)$ as the minimal subfield containing both K and α. We already know that we want a field, so we know that $K(\alpha)$ needs to be closed. This means that selecting any element from the field and then multiplying or adding other elements in the field a finite number of times must result in another element in the field. Multiplying or adding any number of elements from the original field together results in an element in the original field (because our original field is closed), so we can restrict our attention to the element that we are adjoining to the field. How do we represent adding and multiplying a bunch of elements (including some number of multiplications by α)? Well, if we were to group the terms together by the number of times that α is multiplied in each term, we could collapse the resulting form into something that looked exactly like a polynomial that you've plugged α into: $a_n\alpha^n+...+a_1\alpha+a_0$ (as the non-α elements in each term are all in *K*, they can be collapsed into a single element in *K*). This could lead to a useful notation: $K[\alpha]=\{f(\alpha): f(\alpha)\in K[x]\}$. Informally, $K[\alpha]$ is the smallest ring that is closed under addition and multiplication that contains both the field *K* and α.

This covers the "closed" requirement, but sadly we aren't necessarily left with a field after this step: we still need multiplicative inverses. For this requirement, we'll proceed with a ham-fisted approach and just tack on every possible inverse as well. We'll do this by pulling a similar trick as above: $K(\alpha)=\left\{\frac{f(\alpha)}{g(\alpha)}: f(\alpha),g(\alpha)\in K[x],g(\alpha)\neq 0\right\}$. Addition and multiplication in this set is handled the same way as a rational expression: $\frac{f(\alpha)}{g(\alpha)}\frac{h(\alpha)}{j(\alpha)}=\frac{f(\alpha)h(\alpha)}{g(\alpha)j(\alpha)}$ and $\frac{f(\alpha)}{g(\alpha)}+\frac{h(\alpha)}{j(\alpha)}=\frac{f(\alpha)j(\alpha)+h(\alpha)g(\alpha)}{g(\alpha)j(\alpha)}$. This forms a field, but is this the minimal field? Again, assume that we had some "additional" element $\frac{f(\alpha)}{g(\alpha)}$ contained in our construction that is not in the minimal field. We know that $f(\alpha)\neq 0$ and $g(\alpha)\neq 0$ (if $f(\alpha)$ were 0, then $\frac{f(\alpha)}{g(\alpha)}$ would be 0, which is required to be in the field, and if $g(\alpha)$ were 0 then the expression would not have been included in the first place). We know that both $f(\alpha)$ and $g(\alpha)$ must be in the minimal field, by closure. The fact that $g(\alpha)$ is in the minimal field means that $\frac{1}{g(\alpha)}$ is also in the field. Multiply this by $f(\alpha)$, and we get $\frac{f(\alpha)}{g(\alpha)}$, which must also be in the minimal field. This is the element that we had assumed was not in the minimal field, so this is a contradiction. This contradiction implies that our construction must be minimal.

Note that $K[\alpha]\subseteq K(\alpha)$, as $g(x)=1$ is a perfectly fine polynomial to pick for the denominator, and the numerator can still range over all the polynomials in $K[x]$. This inclusion will be important in just a little bit.

Kronecker's Theorem, etc.

We are now going to develop a relationship between $K[x]/\langle m_{\alpha,K}(x)\rangle$ and $K[\alpha]$. We know that $K[x]/\langle m_{\alpha,K}(x)\rangle$ is a field, and we would like to demonstrate that $K[\alpha]$ is also a field. We can't say that the two sets are equal, precisely, because we write elements in them differently. What we really want is a way of saying that the two sets are morally the same: that they are the same other than the way that we write the elements. This sort of relationship is called being "isomorphic" (which is literally translated to meaning that the fields have the "same shape"). In order to designate that two fields A and B are isomorphic, we write $A \approx B$.

There are two basic requirements for two sets to be isomorphic (as fields):

- There must be a map that "translates" the elements of the first set so that they look like elements of the second set, and we must be able to unambiguously translate set elements back (so there must be a bijection between the first and second field.)
- If we add or multiply two elements in the first set and then translate the result into the second set, it is the same as if we translated the two initial elements into the second set, and then added or multiplied there (this bijection must respect addition and multiplication.)

A map that fulfills these requirements is called a field isomorphism. If two sets are isomorphic as fields, and one of them is a field, then the isomorphism indicates that both sets are fields. This is the particular characteristic that we're going to use to show that $K[\alpha]$ is a field.

A Candidate Isomorphism

We'll evidence a particular map, and demonstrate that it fulfills the field isomorphism requirements, which will imply that $K[x] / \langle m_{\alpha,K}(x) \rangle$ and $K[\alpha]$ are isomorphic. Because we know that $K[x] / \langle m_{\alpha,K}(x) \rangle$ is a field, this will in turn imply that $K[\alpha]$ is a field.

The mapping that we'll look at is this: if $f(x) + \langle m_{\alpha,K}(x) \rangle \in K[x] / \langle m_{\alpha,K}(x) \rangle$, the map is $\Psi\left(f(x) + \langle m_{\alpha,K}(x) \rangle \right) = f(\alpha)$ (For those of Mathy persuasion, this is the substitution ring homomorphism, which one can use to prove that these are isomorphic via the first isomorphism theorem).

The Candidate is a Function

The very first thing that we need to consider is whether this is a valid function: is it single valued and well defined? First, is the function single valued? That is, when we input a single value, does a single value come out (this is implicit in the modern definition of "function")?

For the question of single valued, we can note that polynomials are themselves functions, so just plugging in a number (artificial though it may be), will result in a single value. This implies that our map is single valued.

The question of "well defined" is a technical but important one: the domain is a set of equivalence classes, so we need to make sure that two different members of the same equivalence class are dealt with the same way by the mapping. In other words, we've established a rule that allows us to say if two different polynomials are in some sense "the same" (they have the same remainder after being divided by the minimal polynomial). The question is: upon passing two polynomials in the same equivalence class through our map, is the result the same?

Let's take a look: let $\tilde{f}(x)+\langle m_{\alpha,K}(x)\rangle, \hat{f}(x)+\langle m_{\alpha,K}(x)\rangle \in K[x]/\langle m_{\alpha,K}(x)\rangle$ where $\tilde{f}(x)+\langle m_{\alpha,K}(x)\rangle \sim \hat{f}(x)+\langle m_{\alpha,K}(x)\rangle$ (that is, both $\tilde{f}(x)+\langle m_{\alpha,K}(x)\rangle$ and $\hat{f}(x)+\langle m_{\alpha,K}(x)\rangle$ are in the same equivalence class).

We need to verify that $\Psi\left(\tilde{f}(x)+\langle m_{\alpha,K}(x)\rangle\right)=\Psi\left(\hat{f}(x)+\langle m_{\alpha,K}(x)\rangle\right)$. The division algorithm again comes to our aid: if $\tilde{f}(x)+\langle m_{\alpha,K}(x)\rangle \sim \hat{f}(x)+\langle m_{\alpha,K}(x)\rangle$, then it must be the case that $\tilde{f}(x)=m_{\alpha,K}(x)\tilde{g}(x)+r(x)$ and $\hat{f}(x)=m_{\alpha,K}(x)\hat{g}(x)+r(x)$ (the two polynomials are in the same equivalence class; they must have the same remainder $r(x)$). Now, let's see what happens when we apply our map (and plug in α). $\Psi\left(\tilde{f}(x)+\langle m_{\alpha,K}(x)\rangle\right)=\tilde{f}(\alpha)=m_{\alpha,K}(\alpha)\tilde{g}(\alpha)+r(\alpha)=0\tilde{g}(\alpha)+r(\alpha)=r(\alpha)$, and similarly $\Psi\left(\hat{f}(x)+\langle m_{\alpha,K}(x)\rangle\right)=\hat{f}(\alpha)=m_{\alpha,K}(\alpha)\hat{g}(\alpha)+r(\alpha)=0\hat{g}(\alpha)+r(\alpha)=r(\alpha)$. Thus, $\Psi\left(\tilde{f}(x)+\langle m_{\alpha,K}(x)\rangle\right)=\Psi\left(\hat{f}(x)+\langle m_{\alpha,K}(x)\rangle\right)$, and our mapping is well defined, so we have established that it's useful to look into the other properties of this map.

The Candidate is a Bijection

All polynomials in $K[x]$ can be represented using the division algorithm with the divisor $m_{\alpha,K}(x)$. $f(x)\in K[x]$, so we can write $f(x)=m_{\alpha,K}(x)q(x)+r(x)$ with $\deg r(x)<\deg m_{\alpha,K}(x)$. This remainder can be viewed as a member of $K[x]/\langle m_{\alpha,K}(x)\rangle$ (which we write $r(x)+\langle m_{\alpha,K}(x)\rangle$). Plugging in α we again see why Ψ is well defined: $f(\alpha)=m_{\alpha,K}(\alpha)q(\alpha)+r(\alpha)=0q(\alpha)+r(\alpha)=r(\alpha)$.

Recall that each equivalence class has a representative that is the remainder of this particular division, so choosing $r(x)+\langle m_{\alpha,K}(x)\rangle$ (which is certainly in $f(x)\in K[x]/\langle m_{\alpha,K}(x)\rangle$) and plugging in α yields $r(\alpha)=f(\alpha)$. Thus there is an element in $K[x]/\langle m_{\alpha,K}(x)\rangle$ that corresponds to each element in $K[\alpha]$, so this mapping is "onto" (also known as "surjective").

To verify that each element in ${K[x]}\big/{\langle m_{\alpha,K}\rangle}$ corresponds to exactly one element in $K[\alpha]$, we'll assume that we have two elements (possibly in the same equivalence class, possibly in different equivalence classes) in $a(x)+\langle m_{\alpha,K}\rangle, b(x)+\langle m_{\alpha,K}\rangle \in {K[x]}\big/{\langle m_{\alpha,K}\rangle}$ that both map to a particular element in $f(\alpha)\in K[\alpha]$. We want to show that $a(x)+\langle m_{\alpha,K}\rangle \sim b(x)+\langle m_{\alpha,K}\rangle$.

$\Psi\left(a(x)+\langle m_{\alpha,K}\rangle\right)=a(\alpha)=f(\alpha)$ and $\Psi\left(b(x)+\langle m_{\alpha,K}\rangle\right)=b(\alpha)=f(\alpha)$. Again, $f(x)\in K[x]$ so $f(x)=m_{\alpha,K}q(x)+r(x)$, so $f(\alpha)=r(\alpha)$, and $\Psi\left(a(x)+\langle m_{\alpha,K}\rangle\right)=\Psi\left(b(x)+\langle m_{\alpha,K}\rangle\right)=r(\alpha)$. Repeating the above argument, this implies that $b(a)=a(\alpha)=r(\alpha)$ and thus $a(x)+\langle m_{\alpha,K}\rangle \sim b(x)+\langle m_{\alpha,K}\rangle$, so our map is one-to-one (also known as injective).

The Candidate is an Isomorphism

So, now we have a bijection, but we still need to verify that our mapping respects addition and multiplication. This is (happily) pretty easy, once we apply the division algorithm yet again. We'll work with two polynomials, $f(x)$ and $g(x)$ (both in $K[x]$). By the division algorithm, we can represent these two polynomials as we have previously: $f(x)=m_{\alpha,K}(x)q_1(x)+r_1(x)$, $g(x)=m_{\alpha,K}(x)q_2(x)+r_2(x)$.

For addition, we want to verify that

$$\Psi\left(f(x)+\langle m_{\alpha,K}(x)\rangle+g(x)+\langle m_{\alpha,K}(x)\rangle\right)=\Psi\left(f(x)+\langle m_{\alpha,K}(x)\rangle\right)+\Psi\left(g(x)+\langle m_{\alpha,K}(x)\rangle\right)$$

To do this, we'll look at the left and right hand side independently. For the right hand side,

$$\Psi\left(f(x)+\langle m_{\alpha,K}(x)\rangle+g(x)+\langle m_{\alpha,K}(x)\rangle\right)=$$
$$\Psi\left(\left(m_{\alpha,K}(x)q_1(x)+r_1(x)+m_{\alpha,K}(x)q_2(x)+r_2(x)\right)+\langle m_{\alpha,K}(x)\rangle\right)=$$
$$\Psi\left(m_{\alpha,K}(x)\left(q_1(x)+q_2(x)\right)+\left(r_1(x)+r_2(x)\right)+\langle m_{\alpha,K}(x)\rangle\right)=$$
$$m_{\alpha,K}(\alpha)\left(q_1(\alpha)+q_2(\alpha)\right)+\left(r_1(\alpha)+r_2(\alpha)\right)=$$
$$0\left(q_1(\alpha)+q_2(\alpha)\right)+\left(r_1(\alpha)+r_2(\alpha)\right)=$$
$$r_1(\alpha)+r_2(\alpha)$$

Looking at the left hand side,

$$\Psi\left(f(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)+\Psi\left(g(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)=$$
$$\Psi\left(m_{\alpha,K}(x)q_1(x)+r_1(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)+\Psi\left(m_{\alpha,K}(x)q_2(x)+r_2(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)=$$
$$m_{\alpha,K}(\alpha)q_1(\alpha)+r_1(\alpha)+m_{\alpha,K}(\alpha)q_2(\alpha)+r_2(\alpha)=$$
$$0q_1(\alpha)+r_1(\alpha)+0q_2(\alpha)+r_2(\alpha)=$$
$$r_1(\alpha)+r_2(\alpha)$$

They're the same, so we're good with addition.

Multiplication proceeds similarly. We want to show:

$$\Psi\left(\left(f(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)\left(g(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)\right)=\Psi\left(f(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)\Psi\left(g(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)$$

Again, we'll look at the left and right hand side independently. For the right hand side,

$$\Psi\left(\left(f(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)\left(g(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)\right)=$$
$$\Psi\left(f(x)g(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)=$$
$$\Psi\left(\left(m_{\alpha,K}(x)q_1(x)+r_1(x)\right)\left(m_{\alpha,K}(x)q_2(x)+r_2(x)\right)+\left\langle m_{\alpha,K}(x)\right\rangle\right)=$$
$$\Psi\left(\left(m_{\alpha,K}^2(x)q_1(x)q_2(x)+m_{\alpha,K}(x)\left(q_1(x)r_2(x)+q_2(x)r_1(x)\right)+r_1(x)r_2(x)\right)+\left\langle m_{\alpha,K}(x)\right\rangle\right)=$$
$$m_{\alpha,K}^2(\alpha)q_1(\alpha)q_2(\alpha)+m_{\alpha,K}(\alpha)\left(q_1(\alpha)r_2(\alpha)+q_2(\alpha)r_1(\alpha)\right)+r_1(\alpha)r_2(\alpha)=$$
$$0q_1(\alpha)q_2(\alpha)+0\left(q_1(\alpha)r_2(\alpha)+q_2(\alpha)r_1(\alpha)\right)+r_1(\alpha)r_2(\alpha)=$$
$$r_1(\alpha)r_2(\alpha)$$

Looking at the left hand side,

$$\Psi\left(f(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)\Psi\left(g(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)=$$
$$\Psi\left(m_{\alpha,K}(x)q_1(x)+r_1(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)\Psi\left(m_{\alpha,K}(x)q_2(x)+r_2(x)+\left\langle m_{\alpha,K}(x)\right\rangle\right)=$$
$$\left(m_{\alpha,K}(\alpha)q_1(\alpha)+r_1(\alpha)\right)\left(m_{\alpha,K}(\alpha)q_2(\alpha)+r_2(\alpha)\right)=$$
$$\left(0q_1(\alpha)+r_1(\alpha)\right)\left(0q_2(\alpha)+r_2(\alpha)\right)=$$
$$r_1(\alpha)r_2(\alpha)$$

They're the same, so we're good with multiplication.

Consequences of the Isomorphism

We have established that $K[x] \big/ \langle m_{\alpha,K}(x) \rangle$ and $K[\alpha]$ are isomorphic. We know that $K[x] \big/ \langle m_{\alpha,K}(x) \rangle$ is a field, so that implies that $K[\alpha]$ is a field. We know that $K[\alpha] \subseteq K(\alpha)$ (as we established earlier). In addition, we know that $K(\alpha)$ is the minimal field containing both K and α, and that $K[\alpha]$ contains both K and α, so $K(\alpha) \subseteq K[\alpha]$. These two inclusions imply that $K(\alpha) = K[\alpha]$.

Now, let's look at $K[\alpha]$: we know that for all $f(x) \in K[x]$ we can write $f(x) = m_{\alpha,K}(x) q(x) + r(x)$ (by the division algorithm, so $\deg r(x) < \deg m_{\alpha,K}(x)$). Every member of $K[\alpha]$ can be had by plugging α into such a polynomial. As $f(\alpha) = m_{\alpha,K}(\alpha) q(\alpha) + r(\alpha) = 0q(\alpha) + r(\alpha) = r(\alpha)$, this means that every member of $K[\alpha]$ can be had by plugging α into some polynomial with degree smaller than the degree of $m_{\alpha,K}(x)$. What do these polynomials look like?
$K[\alpha] = \{a_0 + a_1\alpha + a_2\alpha^2 + ... + a_{n-1}\alpha^{n-1} : a_k \in K, 0 \le k \le n-1\} = \operatorname{span}_K \{1, \alpha, \alpha^2, ..., \alpha^{n-1}\}$.

As it happens, this list cannot be made smaller, though the proof of this requires a small amount of linear algebra.
If we assume that the set $\{1, \alpha, \alpha^2, ..., \alpha^{n-1}\}$ is linearly dependent (meaning there are extra elements in our spanning set that could be disposed of without reducing the size of the resulting space), then there must be some values for $a_0, a_1, a_2, ..., a_{n-1} \in K$ such that $a_0 + a_1\alpha + a_2\alpha^2 + ... + a_{n-1}\alpha^{n-1} = 0$. We know this cannot be true. If this were true then $g(x) = a_0 + a_1x + a_2x^2 + ... + a_{n-1}x^{n-1}$ would have a root at α and be less than the degree of the minimal polynomial (a contradiction, because the minimal polynomial was selected to have minimal degree!).

So, we have established that (where $n = \deg m_{\alpha,K}(x)$),
$K[x] \big/ \langle m_{\alpha,K} \rangle \approx K[\alpha] = \{a_0 + a_1\alpha + a_2\alpha^2 + ... + a_{n-1}\alpha^{n-1} : a_k \in K, 0 \le k \le n-1\} = K(\alpha)$. This gives us a nice concrete way of dealing with adjoining algebraic numbers to a field, which we can use to great effect in making definitions of the complex numbers.

One might reasonably ask if we can adjoin any number to any field in this way. The answer is, sadly, no. We can use this technique only to adjoin numbers which happened to be roots of polynomials over our field (if there was no such polynomial, there couldn't be a minimal polynomial!). Certain numbers cannot be represented this way over some fields: it is a (non-trivial) fact that both e and π are not roots of any polynomial over the rational numbers, for instance. Such numbers are referred to as transcendental numbers (over our field). Numbers that are roots of a polynomial over our field are called algebraic numbers.

Achtung!
A few words of warning: one might wrongly conclude that because conjugate elements share the same minimal polynomial, that all of the conjugate elements will appear in the field produced by adjoining one of the conjugate elements. We've seen some instances where this happened (in our example, $\mathbb{Q}\left(\sqrt{2}\right)=\mathbb{Q}\left(-\sqrt{2}\right)$ and thus $-\sqrt{2}\in\mathbb{Q}\left(\sqrt{2}\right)$). This is not universally so! If $m_{\alpha_1,K}=m_{\alpha_2,K}$, then it is true that $K\left(\alpha_1\right)=K\left[\alpha_1\right]\approx K[x]/\left\langle m_{\alpha_1,K}\right\rangle=K[x]/\left\langle m_{\alpha_2,K}\right\rangle\approx K\left[\alpha_2\right]=K\left(\alpha_2\right)$, but this only implies that $K\left(\alpha_1\right)\approx K\left(\alpha_2\right)$ (the two fields are isomorphic), NOT THAT THEY ARE EQUAL! Note the other case mentioned in our example:
$m_{\sqrt[3]{2},\mathbb{Q}}=x^3-2$, so $\mathbb{Q}\left(\sqrt[3]{2}\right)=\mathbb{Q}\left[\sqrt[3]{2}\right]=\left\{a_0+a_1\sqrt[3]{2}+a_2\left(\sqrt[3]{2}\right)^2 : a_0,a_1,a_2\in\mathbb{Q}\right\}$. The other elements that share the same minimal polynomial are $\frac{-1}{\left(\sqrt[3]{2}\right)^2}-\frac{\sqrt{3}}{\left(\sqrt[3]{2}\right)^2}i$ and $\frac{-1}{\left(\sqrt[3]{2}\right)^2}+\frac{\sqrt{3}}{\left(\sqrt[3]{2}\right)^2}i$, which are clearly not in $\mathbb{Q}\left(\sqrt[3]{2}\right)$.

2.5 The Field of Complex Numbers

Now that we've proved Kronecker's Theorem and shown some nice ways of adjoining a root to a field, we can use this same formalism to try to extend the real numbers.

Traditionally, we extend the real numbers by adjoining a root of the polynomial $f\left(x\right)=x^2+1$. First, note that this polynomial is irreducible in the real numbers, so we can directly jump to case 3 of Kronecker's Theorem. We'll look at $K=\mathbb{R}[x]/\left\langle x^2+1\right\rangle=\left\{g\left(x\right)+\left\langle x^2+1\right\rangle : g\left(x\right)\in\mathbb{R}[x]\right\}$. We know through our proof of Kronecker's Theorem that this forms a field under the operations that we discussed. Let's look at the operations in more detail. First, note that we get all of the equivalence classes by restricting ourselves to the remainders of polynomial division by x^2+1, so let's just look at those. $z\in K$ implies that we can view *z* as the remainder of this polynomial division (thus *z*, viewed as the remainder of polynomial division, is a polynomial over the real numbers of degree 1 or less). As such, we can view $K=\left\{ax+b : a,b\in\mathbb{R}\right\}$ (to be excruciatingly correct, this is not strict equality, rather it is a complete set of equivalence class representatives).

Let's look into how we can perform addition and multiplication using this field definition:

$w, z \in K$ implies that we can write $w = ax + b$, $z = cx + d$ (for some real values of *a*, *b*, *c*, and *d*). $w + z = (a+c)x + (b+d)$. We didn't change the order of the polynomial, so dividing by $x^2 + 1$ is going to leave the value unchanged, and the exact same value will be the remainder of the division, thus we're still left with $w + z = (a+c)x + (b+d)$.

Now let's look at multiplication: $wz = acx^2 + (ad+bc)x + bd$. Now this division will be productive, so let's do it:

$$\begin{array}{rlllll} & ac & & & & \\ \cline{2-6} x^2+0x+1 \big) & acx^2 & + & (ad+bc)x & + & bd \\ & \underline{acx^2} & \underline{+} & \underline{0x} & \underline{+} & \underline{ac} \\ & & & (ad+bc)x & + & (bd-ac) \end{array}$$

So, the reminder is $wz = (ad+bc)x + (bd-ac)$, which is the equivalence class representative that we're interested in.

This looks like the standard definition of addition and multiplication in the complex numbers, but with *x* instead of the mysterious *i*! Just to verify, let's try to figure out what x^2 is in our field:

$$\begin{array}{rlllll} & 1 & & & & \\ \cline{2-6} x^2+0x+1 \big) & x^2 & + & 0x & + & 0 \\ & \underline{x^2} & \underline{+} & \underline{0x} & \underline{+} & \underline{1} \\ & & & & & -1 \end{array}$$

And there we have it! We now have a field that contains the real values as a sub-field, has a root for $x^2 + 1$, and obeys all the standard conventions for complex number arithmetic. Remember that we claimed that we could view the real numbers as a subfield because arithmetic using the constant polynomials works the same way as real numbers. Thus, it is the constant component of the reduced polynomial that can be thought of as the "real" component, and the coefficient for the x component of the reduced polynomial that can be thought of as the imaginary component.

Def 3: $\mathbb{C} \approx {}^{\mathbb{R}[x]}\!\big/_{\langle x^2+1 \rangle} = \left\{ g(x) + \langle x^2+1 \rangle : g(x) \in \mathbb{R}[x] \right\} = \left\{ ax + b + \langle x^2+1 \rangle : a, b \in \mathbb{R} \right\}$.

As stated earlier, $(ax+b)+(cx+d) = (a+c)x + (b+d)$ and $(ax+b)(cx+d) = (ad+bc)x + (bd-ac)$.

Now that we've shown that our construction fulfills all the characteristics of an extension field, we have the somewhat awkward task of trying to justify why we can think of members of this extension field as numbers. The members of this field look like polynomials, but this is a bit misleading, because we never expect to evaluate the polynomials that form our field. We just called them polynomials because we know how to perform certain operations on polynomials (adding, subtracting, multiplying, and dividing polynomials). We could just as easily extract the underlying operations, and make them a matter of definition:

Def 3a: $\mathbb{C} \approx \left\{(a,b) : a,b \in \mathbb{R}\right\}$. $w,z \in \mathbb{C}$ implies that $w = (a,b)$, $z = (c,d)$ for some real value of *a, b, c,* and *d*. $w + z = (a+c, b+d)$ and $wz = (ac - bd, ad + bc)$.

Note that for this definition, we wrote the "real" component first, and the "imaginary" component second, which is the standard convention (and explains why we have slightly different looking answers here).

Lastly, we get the simplicity of the standard definition "for free". Let's just re-write the definition from 3a, but keep separation using the convention that $i^2 = -1$ (thus in some sense $i = \sqrt{-1}$), and note that it works out the same way:

Def 3b: $\mathbb{C} \approx \left\{a + bi : a,b \in \mathbb{R}\right\}$.

Using the normal arithmetic rules (along with the $i^2 = -1$ rule), we get the rules for addition and multiplication of complex numbers: $w,z \in \mathbb{C}$ implies that $w = a + bi$, $z = c + di$ for some real value of *a, b, c,* and *d*. $w + z = (a+c) + (b+d)i$ and $wz = (ac - bd) + (ad + bc)i$.

Naming this element and providing a rule for arithmetic with it may provide some idea what is next in this regard:

Def 3c: $\mathbb{C} = \mathbb{R}(i) = \mathbb{R}[i] = \left\{a + bi : a,b \in \mathbb{R}\right\} \approx \mathbb{R}[x] \big/ \left\langle m_{i,\mathbb{R}}(x) \right\rangle = \mathbb{R}[x] \big/ \left\langle x^2 + 1 \right\rangle$.

Again, the standard arithmetic rules apply.

So, now we have four (closely related) ways of looking at the complex numbers and the operations involving complex numbers. We have proved (through Kronecker's Theorem) that they form a field, but we never really determined what the multiplicative inverse was in this case. We'll proceed in a similar way as we did in $\mathbb{Q}\left(\sqrt{2}\right)$:

$\frac{1}{(a+bi)} = \frac{1}{(a+bi)}\frac{(a-bi)}{(a-bi)} = \frac{(a-bi)}{(a^2+b^2)} = \frac{a}{(a^2+b^2)} + \frac{-b}{(a^2+b^2)}i$. Our multiplicative inverse $(a+bi)^{-1} = \frac{(a-bi)}{(a^2+b^2)}$.

2.6 All Quadratic Polynomials Factor in $\mathbb{C}$

We have shown that $\mathbb{C} = \mathbb{R}(i)$, and we know that (by construction) $\mathbb{R}(i)$ contains a root of $f(x) = x^2 + 1$ (In fact, it includes both roots, as $f(x) = x^2 + 1 = (x+i)(x-i)$, and $\mathbb{R}(i)$ includes both *i* and -*i*). That's great, but we were looking for the ability to factor every quadratic, not just $x^2 + 1$. Does $\mathbb{C}$ accomplish this task?

Casting our eyes back to the quadratic equation, we see that with this spiff new *i* number, we actually can factor any quadratic polynomial!
$0 = ax^2 + bx + c$, and $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. We only ran into a problem when $b^2 - 4ac < 0$, so let's concentrate on this situation. If $b^2 - 4ac < 0$, then $(-1)(b^2 - 4ac) > 0$. It would be nice if we could just make this change because we want to, but sadly the evil mathematical overlords prevent such activities. Thus, to get the same answer, we need to multiply by something that is equivalent to 1: $[(-1)(-1)](b^2 - 4ac)$, but now we're back in the situation we started with, because it's negative again... Let's regroup and rewrite it: $(-1)[(-1)(b^2 - 4ac)]$. Putting this back into the quadratic equation, we get
$x = \dfrac{-b \pm \sqrt{(-1)[(-1)(b^2 - 4ac)]}}{2a} = \dfrac{-b \pm i\sqrt{[(-1)(b^2 - 4ac)]}}{2a}$, and now we're cooking with grease! $(-1)(b^2 - 4ac) > 0$, so everything is grand, other than the new *i* that we just introduced.

We just saw that every polynomial over the real numbers has roots in $\mathbb{C}$, but it isn't clear that we couldn't get some other (smaller) field by just adjoining the roots of some other particular quadratic to the real numbers. If the quadratic has one or two real roots, then clearly we don't need to adjoin anything to the real numbers at all. For a more interesting case, let's look at a quadratic polynomial with no real roots. Let's say that one of the roots is $a + bi$; Let's just adjoin this one root and see what happens. $\mathbb{R}(a+bi)$ needs to contain all of the real numbers, and the root $a + bi$. This is a subset of $\mathbb{R}(i)$, as $\mathbb{R}(i)$ contains both the real numbers and $a + bi$ (so $\mathbb{R}(a+bi) \subseteq \mathbb{R}(i)$). Now note that we can start with $a + bi$, subtract the real value *a*, and then divide by the real value *b* (b isn't zero in this case, as we assumed that the quadratic polynomial has no real roots), leaving us with *i*. This shows that *i* is also an element of $\mathbb{R}(a+bi)$. By the definition of $\mathbb{R}(i)$, we now know that $\mathbb{R}(i)$ is a subset of $\mathbb{R}(a+bi)$ (so $\mathbb{R}(i) \subseteq \mathbb{R}(a+bi)$). As we have setwise inclusion in both directions, we know that $\mathbb{R}(i) = \mathbb{R}(a+bi)$.

Stepping back, we just found that $\mathbb{R}(i)$ allows us to solve for the roots of all quadratics and that if we adjoined the root of any other quadratic without a solution in the real numbers, we get the same field, $\mathbb{C}$.

3 The Complex Plane (stub)

By looking at Definition 3a, we can see that this is notationally similar to the way that we write point on the Cartesian plane. We can think of the pair (a,b) as a point on a plane in rectangular coordinates. This view of the complex numbers is profoundly useful, and yet it wasn't accepted until after Gauss! This planar view of the complex numbers can help to provide a wonderful geometric flavor for the complex numbers which can aide understanding, but this view should not completely supplant our various field definitions. Points on a plane do *not* form a field. It is only through the rather particular definitions of addition and multiplication that we got the wonderful field structure.

As we can think of the pair as a point (a,b) on a plane in rectangular coordinates, it seems reasonable to inquire if other methods Of course, what we think of in rectangular coordinates can be thought of in polar coordinates, and then converted back to rectangular coordinates.

Notation with cis, etc
modulus, argument

addition/multiplication, and its effect

There is some special insight to be gained by looking at the value that we had to multiply $a+bi$ in order to get a real number. We used $a-bi$, which is called the conjugate of $a+bi$. If $z=a+bi$ this is written $\overline{z}=a-bi$.

3.1 *exp(z) on the Complex Plane*

Euler's formula: Two ways to proceed:
Approach 1:
Derive power series for exp
look at power series rep of sin, cos.
Factor, out pops exp.

Asides: def of sin(z), cos(z).
e is the foundation for periodic functions
connection to forrier analysis
e can be used to represent periodic phenomena!

Implies that exp(ix) is on the unit circle, and that we can get any coordinate with this representation.

Approach 2:
calculus derivation:
z = cos(x) + sin(x) i

and notice that when $x = 0$, $z = 1$. Then differentiate,

$dz/dx = -\sin(x) + \cos(x) i$
$dz/dx = \sin(x) i2 + \cos(x) i$
$dz/dx = [\cos(x) + \sin(x) i]i$
$dz/dx = zi$
$(1/z)dz/dx = i$
$\ln(z) = xi + C$

for some constant C, by indefinite integration. Now use the fact that when $x = 0$, $z = 1$, to conclude that $C = 0$. Thus

$\ln(z) = xi$
$z = exi$
$e^{(xi)} = \cos(x) + \sin(x) i$

Dr. Euler's Magic formula
Not terribly profound once the terms are defined.
dependent on radian angle measure

The Euler helps tremendously in a few ways:
roots of unity (with geometry of solution)

subtleties:
Define log as inverse of exp
Multi-valued function, and branch points
Definition of a^b in complex values (finite, infinite values)
Note that there are many ways of representing the same rectangular coordinate.
In some areas, we can treat all of these as identical
Sadly, sometimes, not so much.

$i^{(-i)}=(e^{pi})^{(1/2)}$
$(-1)^{(-i)}=e^{pi}$

topological view (lift)

Talk about complex conjugate

3.2 Oddities of the Complex Plane

Cauchy-Riemann equation
Cauchy Integration
Liouville's Boundedness Theorem
The connection between Residues and Integration

3.3 Applications of the Complex Plane (stub)

3.3.1 Fundamental Theorem of Algebra (stub)

consequences

- every polynomial over the complex numbers splits
- every polynomial over the reals splits in the complex numbers
  - every polynomial over the reals splits into quadratics and linear terms

3.3.2 Thermodynamics (stub)

Sensor behavior
1-e^(-x) (asymptotic approaches new temperature)

3.3.3 Fluid Dynamics (stub)

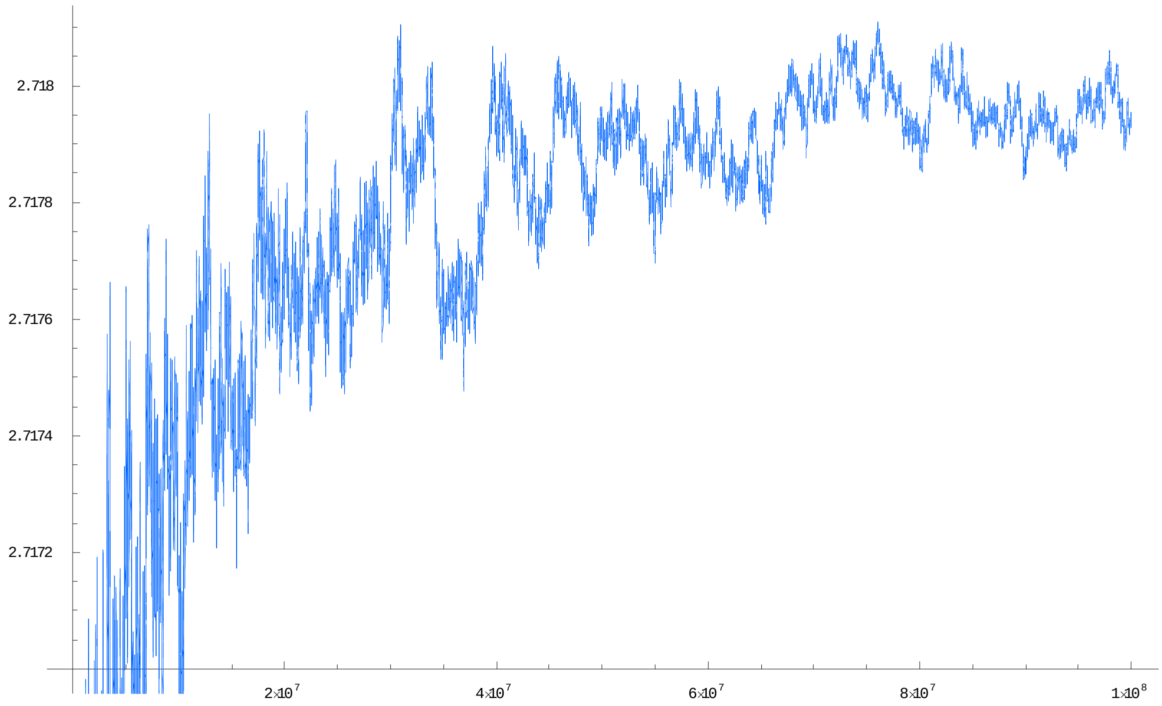
4 Hypergeometric Series and Basic Hypergeometric Series (stub)

5 *Curiouser and Curiouser!*

5.1 The Prime Number Theorem (stub)

Develop via Newman's analytic proof. Prove $\pi(x) \sim \frac{x}{\log(x)}$. Then show def of $\theta(x) = \sum_{p \le x} \log p$. Show that PNT is equivalent to $\lim_{n\to\infty} \frac{\theta(n)}{n} = 1$, so $\lim_{n\to\infty} \frac{1}{n} \sum_{p \le x} \log p = 1$.

Collapse summation to $\lim_{n\to\infty} \frac{1}{n} \log\left(\prod_{p \le n} p\right) = 1$, so $\lim_{n\to\infty} \log\left(\prod_{p \le n} p\right)^{\frac{1}{n}} = 1$ iff $\lim_{n\to\infty}\left(\prod_{p \le n} p\right)^{\frac{1}{n}} = e$



5.2 Counting Words

Q: What are the number of ways of selecting words (order is relevant, no repetition is allowed) from *n* distinct letters (where n is a positive integer)? (i.e., rearrange the n letters into strings from 0 to n letters long)

A: $\lfloor en! \rfloor$

Why? The total number of strings that can be assembled by arranging n letters into a k-length string is $n^{\underline{k}} = (n)(n-1)(n-2)\ldots(n-k+1)$ (called the falling factorial).

As each set of k-length words is disjoint (were the not disjoint, some string would have more than one length), the total number of strings is $f(n) = \sum_{k=0}^{n} n^{\underline{k}} = \sum_{k=0}^{n} \frac{n!}{(n-k)!}$. By reversing the sum, we get $f(n) = \sum_{k=0}^{n} \frac{n!}{k!}$. The *n!* term is invariant in this sum, so we can pull it out: $f(n) = n! \sum_{k=0}^{n} \frac{1}{k!}$. Note that this looks rather like the series expansion for *e*, multiplied by *n!*. Let's see how close they are by subtracting:

$$n!e - f(n) = n!\sum_{k=0}^{\infty} \frac{1}{k!} - n!\sum_{k=0}^{n} \frac{1}{k!} = \sum_{k=n+1}^{\infty} \frac{n!}{k!} = \frac{n!}{(n+1)!} + \frac{n!}{(n+2)!} + \ldots = \frac{1}{(n+1)} + \frac{1}{(n+1)(n+2)} + \ldots$$

so $n!e - f(n) < \frac{1}{(n+1)^1} + \frac{1}{(n+1)^2} + \ldots = \sum_{k=0}^{\infty} \left(\frac{1}{n+1} \right)^k - 1$. Note this last series is geometric, and converges to $\frac{1}{1-\frac{1}{n+1}}-1=\frac{1}{n}$. This implies that $0<n!e-f\left(n\right)<\frac{1}{n}$. For our selection, $n\geq 1$. For *n=1*, $f(1)=2=\lfloor e\rfloor$, so our answer holds. For $n\geq 2$, we have $0<n!e-f\left(n\right)<\frac{1}{2}$, so $f(n)=\lfloor en!\rfloor$. Thus, $f(n)=\lfloor en!\rfloor$ for all positive integer values of *n*.

5.3 The Hat Check Problem

Q: *n* men with hats go to a party and check their hats. At the party they get quite drunk, and as they leave they choose a random hat from the hat check. What is the probability that they have all chosen the incorrect hat?

A: To quote Bryce Jenkin, "$\frac{1}{e}$. Everyone knows that!" More properly, as the number of hat-wearing party goers goes to infinity, the probability approaches $\frac{1}{e}$.

If p is the probability of this occurrence, the answer is ultimately going to be an expression that looks like this: $p = \frac{\text{\# of possibilities in which no one gets the right hat}}{\text{\# of possibilities total}}$.

First, we note that in calculating the number of possible hat rearrangements, we are looking at the number of permutations of *n* distinct items: there are *n!* total ways that the hats can be rearranged. Hey, we have the denominator of our answer! We're half done!

Now, to the numerator: we want the total number of permutations that include no "fixed points", that is the number of permutations where everyone is not matched with their own hat. The name for such permutations is "derangements".

In order to calculate the number of derangements, we're going to have a pleasant aside into an area of math called "combinatorics", or the art of counting without counting. In particular, we're going to look at the very powerful technique called "The Principal of Inclusion/Exclusion", or PIE.

The general problem can be described by our friend, the Venn diagram (see figure 2).

We want to count the total number of things in the universe of possibilities (*U*), which are not in the set *A*, *B* or *C*. We'll keep track of what we've counted in a table that will record the process results up to that round. We want to end up with the table:

Round	U	Just A	Just B	Just C	Just A⋂B	Just B⋂C	Just A⋂C	A⋂B⋂C
?	1	-1	-1	-1	-1	-1	-1	-1

This is read "All the units in *U* are counted, and then we subtracted one for each item in just *A*, one for each item in just *B*, one for each item in just *C*, one for each item just in both *A* and *B*, one for each item just in *B* and *C*, one item just in *A* and *C*, and finally one item in *A* and in *B* and in *C*."

Let's start big, and work our way in.

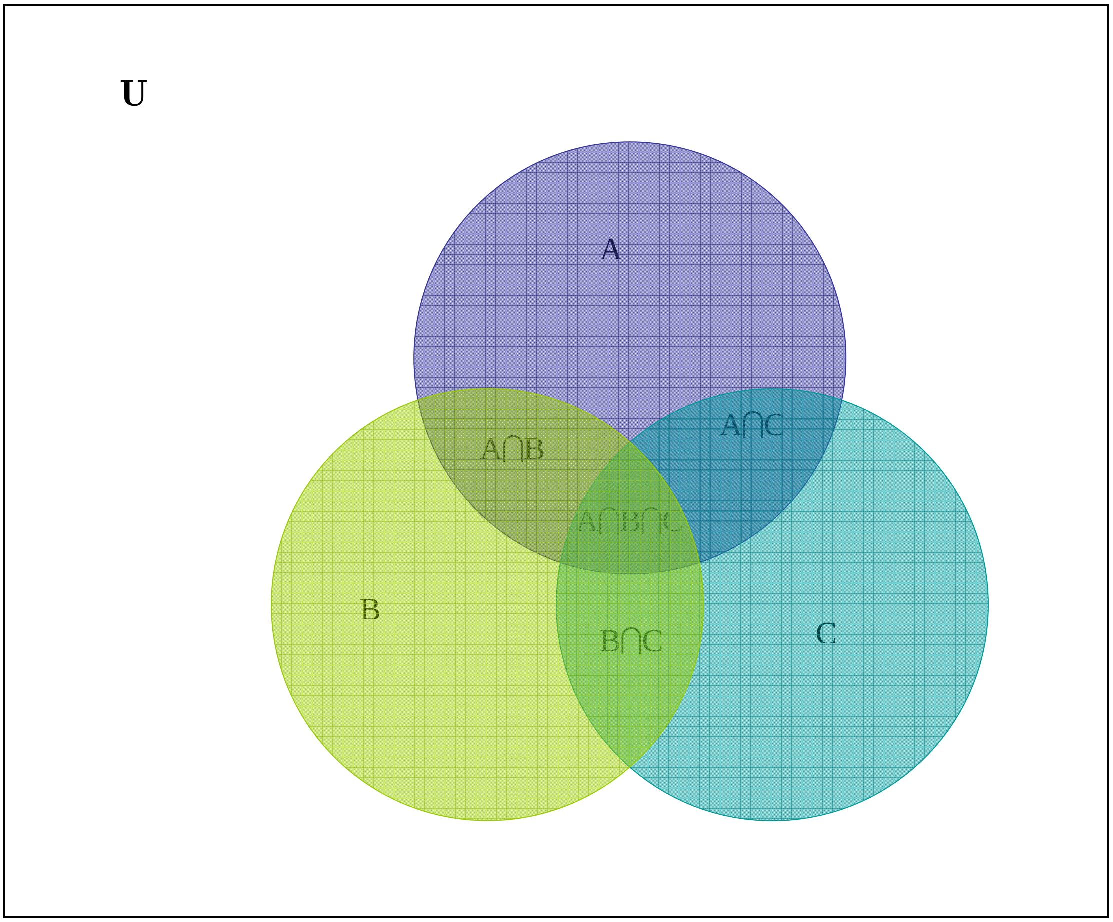


Figure 2

For our first step, let's just take the number of elements in *U*.

Round	*U*	Just *A*	Just *B*	Just *C*	Just *A*⋂*B*	Just *B*⋂*C*	Just *A*⋂*C*	*A*⋂*B*⋂*C*
1	1	0	0	0	0	0	0	0

That's clearly not it, as we want to exclude everything in the sets *A*, *B* and *C*! Let's subtract the number of items in *A*, *B*, and *C*.

Round	*U*	Just *A*	Just *B*	Just *C*	Just *A*⋂*B*	Just *B*⋂*C*	Just *A*⋂*C*	*A*⋂*B*⋂*C*
1	1	0	0	0	0	0	0	0
2	1	-1	-1	-1	-2	-2	-2	-3

Note that we just over-counted the number of items to remove, as there could be items in both *A* and *B* (*A*⋂*B*), in both *B* and *C* (*B*⋂*C*) and in both *A* and *C* (*A*⋂*C*). Now, let's undo our over-counting by adding back the number of items in the sets *A*⋂*B*, *B*⋂*C* and *A*⋂*C*.

Round	U	Just A	Just B	Just C	Just $A \cap B$	Just $B \cap C$	Just $A \cap C$	$A \cap B \cap C$
1	1	0	0	0	0	0	0	0
2	1	-1	-1	-1	-2	-2	-2	-3
3	1	-1	-1	-1	-1	-1	-1	0

That's even better, but now we've over-counted the number of items to add back, because the items that were in *A* and *B* and *C* ($A \cap B \cap C$) were added back three times, but they only should have been added back twice. To correct, we'll re-subtract the number of items in $A \cap B \cap C$.

Round	U	Just A	Just B	Just C	Just $A \cap B$	Just $B \cap C$	Just $A \cap C$	$A \cap B \cap C$
1	1	0	0	0	0	0	0	0
2	1	-1	-1	-1	-2	-2	-2	-3
3	1	-1	-1	-1	-1	-1	-1	0
4	1	-1	-1	-1	-1	-1	-1	-1

And finally, we are victorious! Now, we just need to systematize this process, and stick it in a heinous looking formula to finish the task.

PIE: $$\left|\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n\right| = \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|} \left|\bigcap_{j \in I} A_j\right|$$

Well, that certainly qualifies as a heinous looking formula. Let's try to decode this statement.

From the left hand side, start with $\left|\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n\right|$: in our example, we tried to count every item in the universe (*U*) that was not in the sets *A*, *B* or *C*. In our new notation, we don't just have 3 sets that we want to exclude, but *n* sets, so we label them A_1 through A_n. The bar over each set is the setwise complement, otherwise written $\overline{A}_j = U - A_j$ (or every item in the set *U* other than the elements in the set A_j). The intersection of all the setwise complements gives us all the elements that are in the setwise complement of every one of our sets. In other words $\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n$ is all the elements in *U* that aren't in any of the sets A_1 through A_n. The bars indicate that we are looking for the count of (distinct) elements in the sets, not the sets themselves. So, in summary, the left hand side says "the number of elements that are not in any of the sets A_1 through A_n".

The right hand side has a number of distinct elements that we'll examine from left to right. First, the summand: $\sum_{I \subseteq \{1,\ldots,n\}}$. The presence of the summation implies that we will have a sequence of numbers that we want to add. Because of the way that summation works, the index ($I \subseteq \{1,\ldots,n\}$) should in some sense govern the number added. So, what does $I \subseteq \{1,\ldots,n\}$ mean, anyway? We are using the index variable *I*, which indexes over every possible subset of the set $\{1,2,...,n\}$. By convention, the empty set ($\varnothing$) is a subset of every set, so even if there are 0 sets to be excluded, the summation will at least add one number.

So, how many numbers are we going to add? That's the same as the question "how many distinct subsets are there for an n-set ($\{1,2,...,n\}$)?" Well, that question, at least, is fairly straight forward: Think of every possible set element (the numbers 1 to *n*) as a distinct bit location in an n-bit long bit string; for the purposes of our conversation, let's number our bits from right (bit 1) to left (bit n). Each possible subset of our n-set corresponds to a single binary string (where the bit corresponding to each subset member is a '1', and every bit corresponding to an element not in the subset is a '0'.) This establishes something called a bijection (a surjective, injective function) between the set of three bit strings, and the subsets of the 3-set ($\{1,2,3\}$). For clarity, here is the explicit bijection:

Subset	Bitstring
{}	000
{1}	001
{2}	010
{3}	100
{1,2}	011
{2,3}	110
{1,3}	101
{1,2,3}	111

We know that an n-bit bit-string has 2^n possible settings, so our n-set has 2^n distinct subsets.

So, when we have n sets to exclude from our count, we will index over 2^n sets (and thus add together 2^n numbers). But what are these numbers? That's established by the next two portions of our formula. The first is fairly straight forward, and it just establishes the sign of the number added: $(-1)^{|I|}$. This expression changes the sign of the number to a negative value in the case that there are an odd number of items in the particular subset we're looking at (when *I* contains an odd number of elements), and leaves it as a positive number in the case where there is an even number of elements in our subset (when *I* contains an even number of elements).

The next portion is notationally distressing, but actually fairly straight forward: $\left|\bigcap_{j\in I} A_j\right|$ is just counting the number of common elements in the sets referenced by the current value of *I*. So, if the current value of *I* happened to be $\{1,3,6\}$ then $\left|\bigcap_{j\in I} A_j\right| = |A_1 \cap A_3 \cap A_6|$.

There is one convention to note, however. Remember that $\varnothing$ is a subset, so we'll need to deal with the case where $I = \varnothing$. This is the special case called the "empty intersection"; we'll adopt the convention that any empty intersection produces the entire universal set, *U*. (This is intuitively justified by noting that intersections are all about pruning non-common elements, so if one never prunes anything off, one is left with everything, or the universal set).

So, we're done with the formula, but it isn't very clear how our example in any way corresponds to out formula, but it does! Let's first change notation: let A_1=*A*, A_2=*B* and A_3=*C*. For the index of the summation, we are indexing over subsets of the set $\{1,2,3\}$.

Our PIE formula for this example is this: $\left|\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3\right| = \sum_{I \subseteq \{1,2,3\}} (-1)^{|I|} \left|\bigcap_{j \in I} A_j\right|$.

Refer back to the example, and note that in round 1 we just added all of the elements in *U*. This corresponds to the index $I = \varnothing$, so $(-1)^0 \left|\bigcap_{j \in \varnothing} A_j\right| = |U|$.

In round 2, we subtracted the number of elements in A_1, A_2 and A_3. This corresponds to the three indexes $I = \{1\}$, $I = \{2\}$, and $I = \{3\}$. Note that our formula here is adding a negative value for each of these indexes, as $|I| = 1$ in each case. For $I = \{1\}$, $(-1)^1 \left|\bigcap_{j \in \{1\}} A_j\right| = -|A_1|$, and similarly so for $I = \{2\}$ and $I = \{3\}$.

In round 3, we then added the number of elements shared by any two sets: $|A_1 \cap A_2|$, $|A_2 \cap A_3|$, $|A_1 \cap A_3|$. These correspond to the indexes $I = \{1,2\}$, $I = \{2,3\}$, and $I = \{1,3\}$, respectively. For these subsets, we have an even number of elements in our index (2), so we are not changing the sign. In the case where $I = \{1,2\}$, our formula becomes $(-1)^2 \left|\bigcap_{j \in \{1,2\}} A_j\right| = |A_1 \cap A_2|$, with similar results for the other two indexes of this sort ($I = \{2,3\}$ and $I = \{1,3\}$).

In round 4, we subtracted the number of elements common to them all: $|A_1 \cap A_2 \cap A_3|$. This corresponds to the index $I = \{1,2,3\}$. Once again, our formula works out: we subtracted, because there were an odd number of elements in our index, and our formula for this round is $(-1)^3 \left|\bigcap_{j \in \{1,2,3\}} A_j\right| = -|A_1 \cap A_2 \cap A_3|$.

We have now indexed over the $2^3 = 8$ different subsets of our 3-set ($\{1,2,3\}$), so we're done. Putting it all together, we get:

$$\left|\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3\right| = |U| - |A_1| - |A_2| - |A_3| + |A_1 \cap A_2| + |A_2 \cap A_3| + |A_1 \cap A_3| - |A_1 \cap A_2 \cap A_3|$$

This is exactly what we got last time, so it seems reasonable that PIE may be true, even though it looks a bit strange.

Having informally talked about this result for pages, it might seem that the proof is likely to be hideous, but it's fairly straight forward:

PIE: $$\left|\overline{A}_1 \cap \overline{A}_2 \cap ... \cap \overline{A}_n\right| = \sum_{I \subseteq \{1,...,n\}} (-1)^{|I|} \left|\bigcap_{j \in I} A_j\right|$$

Proof:
Proving this will require a new way of looking at PIE. We've been principally interested in the final answer thus far, but in order to verify that PIE works, we need to follow a particular element and determine how it is accounted for as we perform the summation. We'll choose an arbitrary member of *U* (which we'll call *x*) and follow how it's counted. For this discussion, we'll adopt the terminology "positively counted" if *x* is present in $\bigcap_{j \in I} A_j$ term when *I* has an even number of elements, "negatively counted" if *x* is present in the $\bigcap_{j \in I} A_j$ term when *I* has an odd number of elements and "not counted" if *x* is not in the $\bigcap_{j \in I} A_j$ term at all. We'll want to show that each item in *U* will either be positively counted exactly once, in the case where it is in none of the sets that we are excluding (and this element will contribute to the end answer) or alternately (in the case where the particular item we're following is in at least one of the excluded sets) that the number of times that it is positively counted is equal to the number of times it is negatively counted (in which case this item does not influence the end answer).

Let $x \in U$. We have only two cases that we need to look at:
Case 1: *x* is not in any of the *n* sets A_1 ... A_n. In this case, we want *x* to be positively counted exactly once. Looking at PIE, *x* is going to be positively counted in the instance where the index is $\varnothing$, and then is going to be not counted for every other index. Why? If it were ever counted for a non-empty index then it would have been in at least one of the sets A_1 ... A_n. We know that it isn't (by assumption of the case we're exploring) so *x* is positively counted exactly once.

Case 2: *x* is in some number of sets A_1 ... A_n. In this case, we need to show that ultimately each positive counting is balanced by a negative counting (so that eventually, the sign changes in PIE cause everything to cancel out, leaving us with 0 in the final answer). For the purposes of discussion, let's assume that *x* is a member of exactly *k* ($1 \le k \le n$) of the sets A_1 ... $A_{n..}$ There's no problem with us reordering the sets, so for the purpose of our discussion let's assume that *x* is in the first *k* sets, A_1 ... A_k. How many ways do we have of counting (either positively or negatively) *x* in these *k* sets? Only the intersections involving only the first *k* sets will end up positively or negatively counting our *x* (if the intersection included any sets that were not in the first *k* sets, then *x* would not be in all of the sets, thus their intersection would not include *x* and *x* would then not be counted). There are clearly *k* ways to select one set from our k sets, but what if we wanted to select more than one set? Combinations to the rescue! If we are trying to select *j* sets ($0 \le j \le k$) that all contain our *x* value, there are exactly $\binom{k}{j}$ ways of doing so. Further, this same rule works if *j*>*k*, as in that case $\binom{k}{j}$ becomes 0, which nicely signals the fact that *x* is not being counted. We change the sign of the term depending on the number of sets that we are going to intersect together, so we can deal with both positively counting and negatively counting *x* by changing the sign of our $\binom{k}{j}$ term. Putting all this together, we have the summation $\sum_{j=0}^{n}\binom{k}{j}(-1)^{j}$ which deals with all cases; as j goes from 0 to *k*, the sign alternates appropriately governing positive and negative counting. When *j* is between *k+1* and *n*, the sum automatically does not count *x*. The fact that $\binom{k}{j}$ goes to 0 when *j*>*k* allows us to re-index, leaving us with $\sum_{j=0}^{k}\binom{k}{j}(-1)^{j}$. But wait! This is equivalent to $\sum_{j=0}^{k}\binom{k}{j}(-1)^{j}(1)^{k-j}$, which is the binomial theorem expansion $\sum_{j=0}^{k}\binom{k}{j}(-1)^{j}(1)^{k-j}=(1-1)^{k}=0$.

And thus we have our proof!

So, now we've returned to our main path. Remember, the whole point of developing PIE here was to be able to calculate the number of derangements for n elements. In order to apply PIE, we need to describe each of the sets that we want to exclude. In this case, we'll make *A₁ ... Aₙ*, where a permutation is in *Aₘ* if and only if the *m*th (where $1 \ge m \ge n$) element is a fixed point. We want the total number of derangements, and that's exactly what $\left|\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n\right|$ is. Now that we have that, we return to PIE:

$$\left|\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n\right| = \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|} \left|\bigcap_{j \in I} A_j\right|.$$

Our first step is going to be to classify the subsets by the size, and then combine them with another summation: $\sum_{k=0}^{n} \sum_{\substack{I \subseteq \{1,\ldots,n\},\\ |I|=k}} (-1)^{|I|} \left|\bigcap_{j \in I} A_j\right|$. Now, think about what *k* is in this context: it's the number of elements required to be in each subset. But what does it mean in this context if an element is in a particular subset? It means that this is a fixed point in a permutation, which means that we know precisely where it is in the permutation (it's fixed!). In fact, in determining how many permutations have a certain set of fixed points, we can just discard the number of "places" in the permutation associated with the fixed points, because they are fixed (they never vary, so they can't increase the number of different permutations total). The only variation is from the non-fixed points, and we can calculate the number of possible permutations using only these non-fixed points. The number of permutations of *n* elements is *n!*, but in this case we're fixing *k* of these elements. Now only *n*-*k* of the elements vary, and the total number of permutations possible is $(n-k)!$.

So, we now have

$$\sum_{k=0}^{n}\sum_{\substack{I\subseteq\{1,\ldots,n\},\\|I|=k}}(-1)^k(n-k)! = \sum_{k=0}^{n}(-1)^k(n-k)!\sum_{\substack{I\subseteq\{1,\ldots,n\},\\|I|=k}}1 = \sum_{k=0}^{n}(-1)^k(n-k)!\binom{n}{k}.$$

The next simplification is to note that we can write $\binom{n}{k}=\frac{n!}{k!(n-k)!}$, so we now have

$$\sum_{k=0}^{n}(-1)^k(n-k)!\binom{n}{k} = \sum_{k=0}^{n}(-1)^k(n-k)!\frac{n!}{k!(n-k)!} = n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}.$$

We now have $\left|\bar{A}_1\cap\bar{A}_2\cap\ldots\cap\bar{A}_n\right| = n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}$., which by convention is often referred to as $d_n = n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}$. So, the number of derangements of an n-element set (which by convention is often referred to as d_n) is $d_n = n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}$.

Now, remember that we were ultimately interested in $p=\frac{\text{\# of possibilities in which no one gets the right hat}}{\text{\# of possibilities total}}$, and that we had established that the number of ways total of exchanging *n* hats was *n!*. So,

$p=\frac{n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}}{n!}=\sum_{k=0}^{n}\frac{(-1)^k}{k!}$. Looking back to definition 1d, it seems that this is closely related to the power series representation for $e^{-1}=\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}$. So, we see that as the number of party goers goes to infinity, the probability of everyone getting the wrong hat goes to $\frac{1}{e}$, just as Bryce said.

In fact, we inherit the excellent convergence properties of this power series. At 4 hat-wearing party-goers, we only less than a 2% error, and at 10 hat-wearing party-goers, we have an error of 0.000006%.

6 Symbols and Notation

$f^{-1}(x)$: The compositional inverse of the function f.

$\log_a x$: The logarithm base *a* of *x*.

$\log x$: The logarithm base *e* of *x*.

$\frac{d}{dx}[u]$: The derivative of *u* in terms of *x*.

$f'(x)$: The derivative of *u* (in terms of *x* implicitly in this case).

$\lim\limits_{x\to a} f(x)$: A functional limit.

$\lim\limits_{n\to\infty} f(n)$: A sequence based limit.

$\sum\limits_{k=0}^{n} a_k x^k$: The *n*th partial sum of a power series.

$\binom{n}{k} = \frac{n^{\underline{k}}}{k!}$, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$: The binomial coefficient.

$n^{\underline{k}} = (n)(n-1)(n-2)...(n-k+1)$: Falling factorial.

$k!$: Factorial. $k! = (k)(k-1)(k-1)...(3)(2)(1)$. We'll accept $0! = 1$ by definition.

$\sum\limits_{k=0}^{\infty} a_k x^k$: A power series.

$\int u\,dx$: Indefinite Integration.

$\int_1^x \frac{1}{t}dt$: Definite Integration.

$f^{(k)}(x)$: The *k*th derivative of *f* (implicitly in terms of *x* here).

$|s - x_n| \to 0$ as $n \to \infty$: Equivalent to $\lim\limits_{n\to\infty} x_n = s$.

$\Leftrightarrow$: If and only if. Also written "iff". Can be thought of as "is logically equivalent to".

K, *L*: Fields.

$a \in K$: *a* is chosen from the set *K*.

a^{-1}: The multiplicative inverse of the element *a*.

$a(x) + \langle f(x)\rangle$: A member of the quotient ring $K[x]\big/\langle f(x)\rangle$.

$a(x)^{-1} + \langle f(x)\rangle$: The multiplicative inverse of the element $a(x) + \langle f(x)\rangle$.

$K(\alpha)$: The field K with the element α adjoined to it.

$A = \{\text{set members} : \text{conditions}\}$: Describing the set *A*, which is made up of all of the described set members that fulfill all of the conditions noted.

$K[x]$: The set of (finite degree) polynomials and the zero polynomial with coefficients from K.

$\langle f(x)\rangle$: The ideal formed by $f(x)$. In $K[x]$, $\langle f(x)\rangle = \{f(x)g(x) : g(x) \in K[x]\}$.

${}^{K[x]}\!\big/\!{}_{\langle f(x)\rangle}$: A quotient ring.

$a \sim b$ (context: Abstract Algebra): *a* and *b* are in the same equivalence class (the equivalence relation must be gathered from the context).

$a(n) \sim b(n)$ (context: Analytic Number Theory): $\lim\limits_{n\to\infty}\frac{a(n)}{b(n)} = 1$.

$\hat{C} \subseteq C$: The set $\hat{C}$ is a subset (a proper subset or equal to) of the set C.

$\deg r(x)$: The largest power of x in the polynomial $r(x)$.

$m_{\alpha,K}(x)$: The minimal polynomial for the element α in the field *K*.

$K[\alpha]$: The minimal ring with both K and α.

$A \approx B$: *A* is (field) isomorphic to *B*.

$\overline{z}$: (context: complex analysis) The conjugate of the complex number z.

i: The complex number with the property that $i^2 = -1$.

(a,b): (context intervals): The real numbers between a and b (exclusive).

$[a,b]$: (context intervals): The real numbers between a and b (inclusive).

(a,b): (context coordinate): A point on a two dimensional plane.

$\mathbb{C}$: The complex numbers.

$\mathbb{Q}$: The rational numbers.

$\prod\limits_{p\le x} p$: The product of all primes smaller than *x*.

$\lfloor x \rfloor$: The floor of x (the largest integer less than or equal to x).

$\bigcap$: Setwise intersection.

$\overline{A}$: (context: set theory): The setwise complement of the set A (i.e., $\overline{A} = U - A$ where U is the universal set for this context).

$|A|$: (context: set theory): The number of elements in the set *A*.

$\sum\limits_{I\subseteq\{1,\ldots,n\}} f(I)$: The summation of $f(I)$, where *I* is every possible subset of the set $\{1,\ldots,n\}$.

$\bigcap\limits_{j\in I} A_j$: The intersection of every A_j, where *j* is in the set *I*.

7 References

Rawlings, Don. Lecture Notes for Math 560, Field Theory
Rawlings, Don. Lecture Notes for Math 530, Discrete Mathematics
Stewart, Ian. Galois Theory, 3rd ed. CRC Press, 2004.
Conway, John B. Functions of One Complex Variable I, 2nd ed. Springer-Verlag NY, 1978.
Ahlfors, Lars. Complex Analysis, 3rd ed. McGraw-Hill, 1979
Rudin, Walter. Principles of Mathematical Analysis 3rd ed, McGraw-Hill, 1976
Spivak, Michael; Calculus 3rd ed. Publish or Perish, 1994.
O'Connor, J.J. and Robertson, E. F. The number *e*, http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/e.html
Sebah, Pascal and Gourdon, Xavier. The constant *e* and its computation. http://numbers.computation.free.fr/Constants/E/e.html
Jameson, G. J. O. The Prime Number Theorem.
Knopp, Konrad. Theory and Application of Infinite Series.
Hardy, G.H. and Wright, E. M. An Introduction to the Theory of Numbers.
Apostol, Tom M. Introduction to Analytic Number Theory.
Newman, D.J. "Simple Analytic Proof of the Prime Number Theorem"
Gallian, Joseph. Contemporary Abstract Algebra 5th ed
Ireland, Kenneth and Rosen, Michael. A Classical Introduction to Number Theory, 2nd ed.
Possible: Paul J Nahin, Dr. Euler's Fabulous Formula.
Possible: Paul J Nahin, An imaginary tale.

8 E.E. Hill's Original Paper (stub)

9 Revision History

After Draft 20060723 (0.5.0)

- added the derivation for the quadratic formula
- Finished draft work on field theory section
- added an index, references, and symbol index
- Added forward
- Added bibliography
- General editorial comments
- Made section 1.1 more rigorous
- Edited section 2 for clarity
- Corrected the proof that the non-zero elements in the Quotient Ring in Kronecker's Theorem had multiplicative inverses.
- Revised authorship claim at request of E.E. Hill
- Populated Symbol table
- Integrated editorial comments to sections 2.4 through 2.6

10 Index