Chapter 2

We have seen that even basic concepts like length cannot be adequately described in the system of rational numbers; Given a unit length, some line segments appear to have no measurable extension (i.e. no length) in the rationals. Since this is unacceptable, we are forced to expand our number system to incorporate irrational numbers. And as every irrational number is an infinite sequence of rationals, this, in effect, amounts to a legitimization of infinite descriptions. By closely examining the concept of infinity, we will realize how insignificantly little finite sequences are able to express, it seems that we just can't avoid infinity if we wish to say something important. What is this infinity? Lets introduce it with a paradox (taken from "A First Course in Probability" by Sheldon Ross.)

Suppose that we possess an infinitely large urn and an infinite collection of balls labeled ball number 1, number 2, number 3, and so on. Consider an experiment performed as follows. At 1 minute to 12 P.M., balls numbered 1 through 10 are placed in the urn, and ball number 10 is withdrawn. (Assume the withdrawal takes no time.) At $\frac{1}{2}$ minute to 12 P.M., balls numberered 11 through 20 are placed in the urn, and ball number 20 is withdrawn.

At $\frac{1}{4}$ minute to 12 P.M., balls numbered 21 through 30 are placed in the urn, and ball number 30 is withdrawn. At $\frac{1}{8}$ minute to 12 P.M., balls numbered 31 through 40 are placed in the urn, and ball number 40 is withdrawn, etc. The question of interest is, how many balls are in the urn at 12 P.M.?

The answer to this question is clearly that there is an infinite number of balls in the urn at 12 P.M., since any ball whose number is not of the form $10n$, $n \geq 1$, will have been placed in the urn and will not have been withdrawn before 12 P.M. Hence the problem is solved when the experiment is performed as described.

However, let us now change the experiment and suppose that at 1 minute to 12 P.M. balls numbered 1 through 10 are placed in the urn, and ball number 1 is withdrawn; at $\frac{1}{2}$ minute to 12 P.M., balls numbered 11 through 20 are placed in the urn, and ball number 2 is withdrawn; at $\frac{1}{4}$ minute to 12 P.M., balls numbered 21 through 30 are placed in the urn, and ball number 3 is withdrawn; at $\frac{1}{8}$ minute to 12 P.M., balls numbered 31 through 40 are placed in the urn, and ball number 4 is withdrawn, and so on. For this new experiment how many balls are in the urn at 12 P.M.?

Surprisingly enough, the answer now is that the urn is empty at 12 P.M. For, consider any ball - say, ball number

n. At some time prior to 12 P.M. (in particular, at $\left(\frac{1}{2}\right)^{n-1}$ minutes to 12 P.M.), this ball would have been withdrawn from the urn. Hence for each n, ball number n is not in the urn at 12 P.M; therefore, the urn must be empty at this time.

Thus we see from the preceding discussion that the manner in which the withdrawn balls are selected makes a difference. The reason for that is revealed in the way infinity is defined. First we must learn to count.

The idea behind counting was first formally expressed by Georg Cantor. It is beautifully explained in the article "Infinity" by Hans Hahn:

What do we mean when we say of two finite sets that they consist of equally many things, that they have the same number, that they are equivalent? Obviously nothing more than this, that between the members of the first set and those of the second a correspondence can be effected by which each member of the first set matches exactly a member of the second set, and likewise each member of the second set matches one of the first. A correspondence of this kind is called "reciprocally unique", or simply "one-to-one".

The set of the fingers of the right hand is equivalent to the set of fingers of the left hand, since between the fingers of the right hand and those of the left hand a one-to-one pairing is possible. Such a correspondence is obtained, for instance, when we place the thumb on the thumb, the index finger on the index finger, and so on. But the set of both ears and the set of the fingers of one hand are not equivalent, since in this instance a one-to-one correspondence is obviously impossible; for if we attempt to place the fingers of one hand in correspondence with our ears, no matter how we contrive there will necessarily be some fingers left over to which no ears correspond. Now the number (or cardinal number) of a set is obviously a characteristic that it has in common with all equivalent sets, and by which it distinguishes itself from every set not equivalent to itself. The number 5, for instance, is the characteristic which all sets equivalent to the set of the fingers of one hand have in common, and which distinguishes them from all other sets.

In summary, two sets A and B are equivalent if there is a one-to-one correspondence between them. That is, A and B are equivalent if there exists some function $f: A \to B$ that is both one-to-one and onto. For convenience we may occasionally write $A \sim B$ in place of the phrase "A is equivalent to B". Please note that the relation "is equivalent to" is an equivalence relation. In other words "$\sim$" satisfies the following three properties:

1. It is reflexive: $A \sim A$.
2. It is symmetric: If $A \sim B$, then $B \sim A$.
3. It is transitive: If $A \sim B$ and $B \sim C$, then $A \sim C$.

We are now ready to define infinite sets.

Def: For any positive integer n, let J_n be the set whose elements are the integers $1, 2, \ldots, n$; let J be the set consisting of all positive integers. For any set A, we say:

(a) A is finite if $A \sim J_n$ for some n
(the empty set is also considered to be finite).

(b) A is infinite if A is not finite.

(c) A is countable if $A \sim J$

(d) A is uncountable if A is neither finite nor countable.

(e) A is at most countable if A is finite or countable.

Ex. (a) Let $\mathbb{Z}$ be the set of integers. Then $\mathbb{Z}$ is countable. To see this, define $f: \mathbb{Z} \to J$ by $f(n) = 2n$ if $n \geq 1$ and $f(n) = -2n+1$ if $n \leq 0$. The positive integers in $\mathbb{Z}$ are mapped to the even numbers in J, while 0 and the negative integers in $\mathbb{Z}$ are mapped to the odd numbers in J. That f is both one-to-one and onto is easy to check. Notice that $\mathbb{Z}$ is equivalent to a proper subset of itself! This is typical of infinite sets.

(b) $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$. A quick proof is supplied by the fundamental theorem of arithmetic: Each positive integer $k \in \mathbb{N}$ can be uniquely written as $k = 2^{m-1}(2n-1)$ for some $m, n \in \mathbb{N}$. Define $f: \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by $f(m,n) = 2^{m-1}(2n-1)$. That f is both one-to-one and onto is obvious.

(c) $\mathbb{R} \sim (-\frac{\pi}{2}, \frac{\pi}{2})$. Define $f: \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ by $f(x) = \tan^{-1}(x)$. Recall from calculus that f is a strictly increasing (hence one-to-one) function from $\mathbb{R}$ onto $(-\frac{\pi}{2}, \frac{\pi}{2})$.

Remark: Please observe that the set J, as defined in "Principles of mathematical analysis", is more commonly known as the set $\mathbb{N}$. We will use the lables "J" and "$\mathbb{N}$" interchangeably when we need to describe the set of natural numbers.

Def: By a sequence, we mean a function f defined on the set $\mathbb{N}$ of all positive integers. If $f(n) = x_n$, for $n \in \mathbb{N}$, it is customary to denote the sequence f by the symbol $\{x_n\}$, or sometimes by $x_1, x_2, x_3, \ldots$ The values of f, that is, the elements x_n, are called the terms of the sequence. If A is a set and if $x_n \in A$ for all $n \in \mathbb{N}$, then $\{x_n\}$ is said to be a sequence in A, or a sequence of elements of A. Note that the terms $x_1, x_2, x_3, \ldots$ of a sequence need not be distinct. Since every countable set is the range of a 1-1 function defined on $\mathbb{N}$, we may regard every countable set as the range of a sequence of distinct terms. Speaking more loosely, we may say that the elements of any countable set can be "arranged in a sequence". Sometimes it is convenient to replace $\mathbb{N}$ in this definition by the set of all nonnegative integers, i.e. to start with 0 rather than with 1.

Thm: Every infinite subset of a countable set A is countable.

Proof: Suppose $E \subset A$ and E is infinite. Arrange the elements x of A in a sequence $\{x_n\}$ of distinct elements. Construct a sequence $\{n_k\}$ as follows:
Let n_1 be the smallest positive integer such that $x_{n_1} \in E$. Having chosen $n_1, \dots, n_{k-1}$ $(k = 2, 3, 4, \dots)$, let n_k be the smallest integer greater than n_{k-1} such that $x_{n_k} \in E$. Putting $f(k) = x_{n_k}$ $(k = 1, 2, 3, \dots)$, we obtain a 1-1 correspondence between E and $\mathbb{N}$.

Def: Let A and B be sets. Then $A \setminus B = \{x \in A : x \notin B\}$.

Thm: Every infinite set has a countable subset.

Proof: Let A be an infinite set. Then $A \neq \emptyset$, because $\emptyset$ is finite. Let $x_1 \in A$ be any element of A. Then $A \setminus \{x_1\} \neq \emptyset$ (otherwise $A = \{x_1\}$ and A is finite). Pick $x_2 \in A \setminus \{x_1\}$ to be any element of $A \setminus \{x_1\}$.
Having chosen $x_1, \dots, x_{n-1}$, observe that $A \setminus \{x_1, \dots, x_{n-1}\} \neq \emptyset$ (otherwise $A = \{x_1, \dots, x_{n-1}\}$, making A finite). Hence we are free to select $x_n \in A \setminus \{x_1, \dots, x_{n-1}\}$. Let $E = \{x_n\} \subset A$. Then E is countable, which proves the theorem

The last theorem shows that a countable infinity is the smallest type of infinity; No uncountable set can be a subset of a countable set, while every infinite set has a countable subset.

To motivate our next several results, we present a second proof that $\mathbb{N}\times\mathbb{N}$ is equivalent to $\mathbb{N}$.

Thm: $\mathbb{N}\times\mathbb{N}\sim\mathbb{N}$.

Proof: Arrange $\mathbb{N}\times\mathbb{N}$ in a matrix:

$$\begin{array}{ccccc}
(1,1) & (1,2) & (1,3) & (1,4) & \cdots \\
{\scriptstyle (1)} & {\scriptstyle (3)} & {\scriptstyle (6)} & {\scriptstyle (10)} & \\
(2,1) & (2,2) & (2,3) & \cdots & \\
{\scriptstyle (2)} & {\scriptstyle (5)} & {\scriptstyle (9)} & & \\
(3,1) & (3,2) & \vdots & & \\
{\scriptstyle (4)} & {\scriptstyle (8)} & & & \\
(4,1) & \vdots & & & \\
{\scriptstyle (7)} & & & &
\end{array}$$

The arrows and number marks indicate the order in which we will count the elements of $\mathbb{N}\times\mathbb{N}$. Each diagonal that is traced by the arrows contains all ordered pairs whose components add up to the same number; The components on the first diagonal add up to 2. Those of the second diagonal add up to 3 and so on.

Notice also that the first diagonal contains one element, the second diagonal contains two elements, the third diagonal contains three elements etc. These observations allow us to construct a 1-1 and onto function $f: \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ explicitly:

you should verify that f is given by

$$f(m,n) = \frac{(m+n-2)(m+n-1)}{2} + n.$$

The above theorem gives us a ton of new information. For example:

Thm: The countable union of countable sets is countable; that is, if A_i is countable for $i = 1, 2, 3, \ldots$, then $\bigcup_{i=1}^{\infty} A_i$ is countable.

Proof: Since each A_i is countable, we can arrange their elements collectively in a matrix:

$$\begin{array}{llll}
A_1: & a_{1,1} & a_{2,2} & a_{1,3} \ldots \\
A_2: & a_{2,1} & a_{2,2} & a_{2,3} \ldots \\
A_3: & a_{3,1} & a_{3,2} & a_{3,3} \ldots \\
\vdots & & &
\end{array}$$

and so $\bigcup_{i=1}^{\infty} A_i$ is the range of a map on $\mathbb{N} \times \mathbb{N}$ (How?) That is, $\bigcup_{i=1}^{\infty} A_i$ is equivalent to $\mathbb{N} \times \mathbb{N}$ and hence to $\mathbb{N}$.

Note that the proof of the above theorem can be used to show that, given countable sets A and B, the set $A \times B$ is countable (How?).

Corollary: $\mathbb{Q}$ is countable. (Why?)

Recall that between any two real numbers there is a rational number. This means, in fact, that between any two real numbers, there are infinitely many rational numbers (why?). Surprizingly, $\mathbb{N}$ is as large as $\mathbb{Q}$ even though $\mathbb{N} \subset \mathbb{Q}$ and there are infinitely many rationals between any two natural numbers.

Perhaps you think that all infinities are alike and that $\mathbb{N}$ and $\mathbb{Q}$ are of the same size is a no brainer. Prepare to be shocked.

Thm: $\mathbb{R}$ is uncountable.

Proof: To begin, first note that it is enough to show that $\mathbb{R}$ has an uncountable subset (Why?). Thus, it is enough to show that $(0,1)$ is uncountable. To accomplish this we will show that any countable subset of $(0,1)$ is proper.

Given any sequence $\{a_n\}$ in $(0,1)$, we construct an element x in $(0,1)$ with $x \neq a_n$ for any n. We begin by listing the decimal expansions of the a_n; for example:

$$a_1 = 0.\boxed{3}\ 1\ 5\ 7\ 2 \ldots$$
$$a_2 = 0.\ 0\ \boxed{4}\ 2\ 6\ 8 \ldots$$
$$a_3 = 0.\ 9\ 1\ \boxed{5}\ 3\ 6 \ldots$$
$$a_4 = 0.\ 7\ 5\ 9\ \boxed{9}\ 9 \ldots$$

(If any a_n has two representations, just use the infinite one)

Now let $x = 0.53335 3\ldots$, where the n^{th} digit in the expansion for x is taken to be 3, unless a_n happens to have 3 as its n^{th} digit, in which case we take 5. (This is why we highlighted the n^{th} digit in the expansion of a_n. The choices 3 and 5 are more or less arbitrary here – we just wanted to avoid the troublesome digits 0 and 9.) Then, the decimal representation of x is unique because it does not end in all 0s or all 9s, and $x \neq a_n$ for any n because the decimal expansions for x and a_n differ in the n^{th} place. Thus we have shown that $\{a_n\}$ is a proper subset of $(0,1)$ and hence that $(0,1)$ is uncountable.

The proof that we have just produced is known as Cantor's diagonalization method. It gives insight into the difference between countable and uncountable sets. To understand this difference intuitively, assume that you are a thorough going bureaucrat working at the DMV in a country with infinitely many citizens. If the number of citizens is countable, you will be able to queue all the visitors to the DMV office in a line so that each visitors is serviced eventually. You just have to be clever enough. If, on the other hand, the number of citizens is uncountable, the quantity of visitors can grow to be so large that you cannot even arrange them in waiting order, no matter how you contrive.

Corollary: $\mathbb{R} \setminus \mathbb{Q}$, the set of irrational numbers, is uncountable. (Why?)

So far, we know that the smallest infinite set must be countable. Larger infinite sets are uncountable. Must all uncountable sets be of the same size? We now demonstrate that the answer is no - we can always build bigger and bigger sets.

Given a set A, we write $\mathcal{P}(A)$ for the power set of A - the set of all subsets of A. Now A is clearly equivalent to a subset of $\mathcal{P}(A)$ (namely, the collection of all singletons $\{a\}$, where $a \in A$) but, as it happens, $\mathcal{P}(A)$ is always "bigger" than A:

Cantor's thm: No map $F: A \to \mathcal{P}(A)$ can be onto

Proof: Given any function $F: A \to \mathcal{P}(A)$, define $S_F = \{x \in A : x \notin F(x)\}$, then S_F is a (possibly empty) subset of A. In particular, $S_F \in \mathcal{P}(A)$.

If F is onto, there must be some element $y \in A$ such that $F(y) = S_F$. We now show that no such y can exist: If $y \in F(y)$, then $y \notin S_F$ (why?). On the other hand, if $y \notin F(y)$, then, by definition of S_F, $y \in S_F$. In either case, $F(y) \neq S_F$ for any $y \in A$. Thus F is not onto as desired.

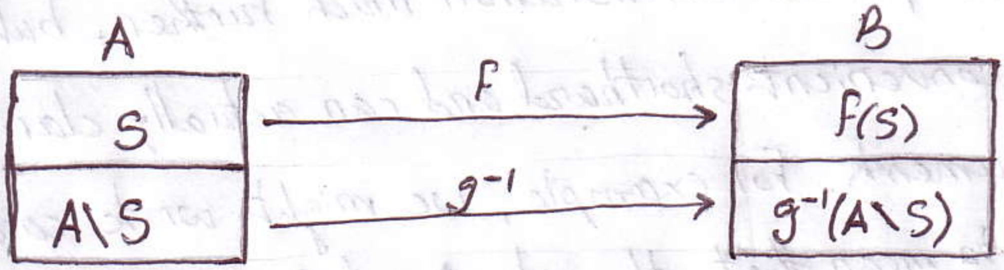
While we won't take the time to fully justify the notation, each set has a cardinal number assigned to it, written card(A) and read "the cardinality of A", that uniquely specifies the number of elements of A. For finite sets the cardinality is literally the number of elements, as in $\text{card}\{1, \ldots, n\} = n$. For countably infinite sets we use the cardinal $\aleph_0$ (read "aleph-naught"), as in $\text{card}(\mathbb{N}) = \aleph_0$. And for $\mathbb{R}$ we write $\text{card}(\mathbb{R}) = c$ (for "continuum").

We will not pursue this notation much further, but it does provide a convenient shorthand and can actually clarify certain arguments. For example, we might write card(A) = = card(B) to mean that the sets A and B are equivalent. And we might use the formula card(A) ≤ card(B) to mean that there is a one-to-one map $f: A \to B$ from A into B. (Why is this a good choice?) But this raises the question of whether the order that we have imposed on cardinal numbers is reasonable. In other words, if card(A) ≤ card(B) and card(B) ≤ card(A) both hold, is it the case that card(A) = card(B)? The answer is "yes" and is given in the following celebrated theorem

F. Bernstein's thm: Let A and B be nonempty sets. If there exist a one-to-one map $f: A \to B$, from A into B, and a one-to-one map $g: B \to A$, from B into A, then there is a map $h: A \to B$ that is both one-to-one and onto.

Proof: We would like to find a set S that will allow us to define $h: A \to B$ as a piece-wise function

$$h(x) = \begin{cases} f(x) & \text{if } x \in S \\ g^{-1}(x) & \text{if } x \in A \setminus S \end{cases}$$



What conditions must the set S satisfy? Since h must be onto B, we must have $B = f(S) \cup g^{-1}(A\setminus S)$ or, equivalently, $A\setminus S = g(B\setminus f(S))$. The last equation may be converted to $S = A\setminus g(B\setminus f(S))$ (How?).
Define $H: \mathcal{P}(A) \to \mathcal{P}(A)$ by $H(S) = A\setminus g(B\setminus f(S))$. We thus have to find a solution to the "fixed point" equation

$$S = H(S).$$

To do this, observe that

i) H is increasing:
Suppose $S \subset T$, then $f(S) \subset f(T)$. Consequently, $B\setminus f(S) \supset B\setminus f(T)$, $g(B\setminus f(S)) \supset g(B\setminus f(T))$, and $A\setminus g(B\setminus f(S)) \subset A\setminus g(B\setminus f(T))$ (Why?) Thus $H(S) \subset H(T)$.

ii) Let $\mathcal{D} = \{S \in \mathcal{P}(A) : S \subset H(S)\}$ Then $\emptyset \in \mathcal{D}$ and $\mathcal{D}$ is not empty.

Let $S^* = \bigcup_{S\in\mathcal{D}} S$, then $S^* \subset H(S^*)$. To see this, observe that for any $S \in \mathcal{D}$, $S \subset S^*$, and $S \subset H(S)$. Since H is increasing, it follows that $H(S) \subset H(S^*)$. Thus $S \subset H(S) \subset H(S^*)$ for all $S \in \mathcal{D}$. Hence $\bigcup_{S \in \mathcal{D}} S \subset H(S^*)$ (Why?)

Notice now that $H(S^*) \subset H(H(S^*))$ (Why?) Thus $H(S^*) \in \mathcal{D}$. It follows that $S^* = H(S^*)$. S^* is therefore the desired set.

Remarkable, right? To help you appreciate how truly incredible Bernstein's result really is, consider the following example.

Ex. Let $\mathbb{R}^\infty$ be the set of all real-valued sequences. That is, if $x \in \mathbb{R}^\infty$, then $x = (x_1, x_2, \dots, x_n, \dots)$ where each $x_i \in \mathbb{R}$. Then $\mathbb{R}^\infty \sim (0,1)$.

To show this, first observe that $\mathbb{R}^\infty \sim (0,1)^\infty$ (Define $f: \mathbb{R}^\infty \to (0,1)^\infty$ by $f(x_1, x_2, \dots) = \left(\frac{\tan^{-1}(x_1) + \frac{\pi}{2}}{\pi}, \frac{\tan^{-1}(x_2) + \frac{\pi}{2}}{\pi}, \dots\right)$)

Thus, it is enough to show that $(0,1)$ is equivalent to $(0,1)^\infty$ - the set of all sequences $\{x_n\}$ with $x_n \in (0,1)$.

To do this, observe that $f:(0,1)\to(0,1)^\infty$ given by $f(x)=(x,0,0,\ldots)$ is a one-to-one map from $(0,1)$ into $(0,1)^\infty$. Thus

$$\operatorname{card}(0,1) \leq \operatorname{card}(0,1)^\infty \qquad (1)$$

To prove the other direction, let $x\in(0,1)^\infty$. Then $x=(x_1,x_2,x_3,\ldots,x_n,\ldots)$, where $x_n\in(0,1)$ for all $n\in\mathbb{N}$. Represent each x_n by its unique infinite decimal expansion $x_n = 0.x_{n1}\,x_{n2}\,x_{n3}\ldots$, let p_n be the n^{th} prime, and define $g:(0,1)^\infty\to(0,1)$ by $g(x)=0.y_1y_2y_3\ldots$ where

$$y_k = \begin{cases} x_{ni} & \text{if } k=p_n^i \\ 0 & \text{otherwise}\end{cases}$$

Then g is one-to-one. In particular,

$$\operatorname{card}(0,1)^\infty \leq \operatorname{card}(0,1) \qquad (2)$$

From conclusions (1) and (2) and Bernstein's theorem, we may conclude that $\mathbb{R}^\infty \sim (0,1)^\infty \sim (0,1)$

Earlier we have demonstrated that the real numbers are uncountable. In fact, the size of the real numbers was the only uncountable infinity of practical importance that has been discussed.

Recall the purpose that the real numbers were designed to serve: The field of rational numbers lacks the least-upper-bound property. As a number system, the field of rationals is consequently too small to relate the length of one object to that of another.

The concept of length truly becomes universal only when measurements are performed in the smallest field that contains the rational numbers and has the least-upper-bound property. The construction in the appendix to Chapter 1 of Rudin's "Principles of Mathematical Analysis" shows that this field is the field of real numbers.

But how much can be described with rational numbers alone? To phrase this question a little differently, consider two line segments. What is the probability that the length of one segment is commensurable with that of the other segment?

Thm: The rational numbers have measure 0 (i.e. occupy no space) on the real number line.

Proof: Since the field of rationals $\mathbb{Q}$ is a countable set, we can list all of its elements in a sequence $\{x_n\}$. We will show that $\mathbb{Q}$ has measure 0 by proving that for any $\epsilon > 0$, there is a collection of open intervals which cover $\mathbb{Q}$ and whose combined length is less than ϵ. To do this, for each $x_n \in \mathbb{Q}$ define I_n by $I_n = (x_n - \frac{\epsilon}{2^{n+1}}, x_n + \frac{\epsilon}{2^{n+1}})$. In other words, I_n is just an interval of length $L(I_n) = \frac{\epsilon}{2^n}$ centered at x_n. Clearly $\mathbb{Q} \subset \bigcup_{n=1}^{\infty} I_n$

Now $L(\bigcup_{n=1}^{\infty} I_n) \leq \sum_{n=1}^{\infty} L(I_n) = \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon \sum_{n=1}^{\infty} \frac{1}{2^n} = \epsilon.$

The above theorem can be interpreted as saying that the likelyhood of selecting a rational number at random in the set of real numbers is 0. To put it in more colorful terms, having selected one object, the chance that another randomly selected object can be described in terms of the first is 0.

Notice that each rational number $\frac{m}{n}$ is a root of the polynomial $p(x) = nx - m$. Also, observe that the n^{th} root of any prime number is a solution to the polynomial equation $x^n - p = 0$. Indeed, the existance of irrational numbers is first braught to light when one considers algebraic relationships between numbers.

Def: A number $x \in \mathbb{C}$ is said to be algebraic if there exist integers $a_0, a_1, \ldots, a_n \in \mathbb{Z}$ such that

$$a_0 + a_1 x + \ldots + a_n x^n = 0$$

Are all or most complex numbers algebraic? The following theorem answers this question with an emphatic no!

Thm: The set of all algebraic numbers is countable.

Proof: Let A_n be the set of all polynomials of degree n with integer coefficients. The map
$a_0 + a_1 x + \ldots + a_n x^n \mapsto (a_0, a_1, \ldots, a_n)$ shows that $A_n \sim \mathbb{Z}^{n+1}$. In particular, A_n is countable (Why?). Now the set of all polynomials with integer coefficients can

be written as the countably infinite union $A = \bigcup_{n=1}^{\infty} A_n$, which must therefore be countable (Why?)

Thus, each polynomial in A can be assigned a natural number that uniquely identifies it.

Let $k \in \mathbb{N}$ be the unique positive integer corresponding to $p(x) = a_0 + a_1 x + \ldots + a_n x^n$. Observe that this polynomial can have at most n distinct complex roots. We can arrange these roots in lexicographic order from smallest to largest and associate $k.1$ with the smallest root of p, $k.01$ with the next smallest root of p, $k.001$ with the third smallest root etc. Clearly, each algebraic number is thus paired up with at least one rational number. This implies that algebraic numbers are countable.

Notice that all countable sets have measure 0 in $\mathbb{R}$ or $\mathbb{C}$ (why?). Thus, the probability that a number is algebraic is 0.

Def: Any complex number that is not algebraic is called transcendental.

It appears that almost all numbers are transcendental, even though we are hard-pressed to name even one.

Having concluded our introductory discussion of infinity, we can now clarify the strange discrepancy between the two seemingly identical procedures of the urn experiment.

To do this, imagine that all the balls are initially inside the urn. The first experiment can then be viewed as asking to remove ball number 10 in the first step, ball number 20 in the second step, ball number 30 in the third step, and so on. That is, the first experiment exhibits a 1-1 correspondence between the set of all natural numbers and the subset of all positive integers that are divisible by 10:

Step #:	1	2	3 ...	n ...
Ball #:	10	20	30 ...	$10n$ ...

Since this subset is proper, we can conclude that the urn is not empty at 12 P.M.

The second experiment requires the removal of the first ball in step 1, the second ball in step 2, and so on. In particular, the second experiment exhibits a 1-1 correspondence between $\mathbb{N}$ and itself.

Step #: $1 \quad 2 \quad 3 \; ... \; n \; ...$

Ball #: $1 \quad 2 \quad 3 \; ... \; n \; ...$

The reason for the apparent similarity of the two experiments: Infinite sets have proper subsets of the same size. The part is as big as the whole.

Metrics and Norms

"In the beginning there were operations - hundreds of them - limits, derivatives, integrals, sums; all of the many operations on functions, sequences, sets, vectors, matrices, and whatever else you might have encountered in calculus. The hallmark of twentieth-century mathematics is that we now view these operations as functions defined on entire collections of 'abstract' objects rather than as specific actions taken on individual objects, one at a time.

Early examples of this type of abstraction appeared in 1906 in Fréchet's thesis, "Sur quelques points du calcul functionnel", in which he introduced a notion of distance defined on abstract sets of points. In particular, Fréchet considered the collection $C[0,1]$, consisting of all continuous real-valued functions defined on the closed interval $[0,1]$, where we measure the distance between two functions by taking the maximum vertical distance between their graphs; that is, $\mathrm{dist}(f,g) = \max_{0 \le t \le 1} |f(t) - g(t)|$. Given a notion of distance between elements of $C[0,1]$, it makes sense to ask questions like: Is integration continuous? That is, are the numbers $\int_0^1 f(t)\,dt$ and $\int_0^1 g(t)\,dt$ "close" whenever f and g are "close"? " (Taken from N.L. Carother's "Real Analysis")

Given a set M, how might we define a distance function on M? What would we want a "reasonable" distance to do? Certainly we would want our distance to be defined and nonnegative for any pair of points in M.

Let's start there: Let $d: M \times M \to [0, \infty)$ be a nonnegative, real-valued function defined on all pairs of elements from M. We would probably expect to have $d(x, x) = 0$ for any $x \in M$. And $d(x, y) = 0$ should mean that $x = y$. We would most likely want our distance to also satisfy $d(x, y) = d(y, x)$ for all pairs of points $x, y \in M$. Anything else? Well, in the hope of preserving at least a bit of the geometry granted by the familiar distances in $\mathbb{R}$ and $\mathbb{R}^n$, we might also require one last property. The distance function should satisfy the triangle inequality is the embodiment of that old saw, "The shortest distance between two points is a straight line." This timid little inequality will turn out to be immensely valuable.

Def: A function d on $M \times M$ satisfying the following properties is called a metric on M.

(i) $0 \le d(x, y) < \infty$ for all pairs $x, y \in M$

(ii) $d(x, y) = 0$ if and only if $x = y$

(iii) $d(x, y) = d(y, x)$ for all pairs $x, y \in M$

(iv) $d(x, y) \le d(x, z) + d(z, y)$ for all $x, y, z \in M$.

The couple (M, d), consisting of a set M together with a metric d defined on M, is called a metric space.

Ex. (a) Every set M admits at least one metric. For example, check that the function defined by $d(x,y) = 1$ for any $x \neq y$ in M, and $d(x,y) = 0$ for all $x \in M$, is a metric. This mundane, but always available, metric is called the discrete metric on M. A set supplied with its discrete metric will be called a discrete space.

(Try to visualize what it would be like to live in a discrete space. Is there a notion of speed? How about accelleration?)

(b) An important example is the real line $\mathbb{R}$ together with its usual metric $d(a,b) = |a-b|$. Any time we refer to $\mathbb{R}$ without explicitly naming a metric, the absolute value metric is always understood to be the one that we have in mind.

(c) Any subset of a metric space is again a metric space in a very natural way. If d is a metric on M, and if A is a subset of M, then $d(x,y)$ is defined for any pair of points $x, y \in A$. Moreover, the restriction of d to $A \times A$ obviously still satisfies properties (i)–(iv).

That is, the metric that is defined on M automatically defines a metric on A by restriction. We will even use the same letter d and simply refer to the metric space (A, d). Of particular interest in this regard is that $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R} \setminus \mathbb{Q}$ each come already supplied with a natural metric, namely, the restriction of the usual metric on $\mathbb{R}$. In each case, we will refer to this restriction as the usual metric

How can we enrich our arsenal of metric functions? To start, suppose that d is a metric on M and $f: M \to M$ is a one-to-one and onto function. Then $\rho: M \times M \to \mathbb{R}$ defined by $\rho(x,y) = d(f(x), f(y))$ is also a metric on M as you should verify.

Also if $g: M \to \mathbb{R}$ is a 1-1 function (not necessarily onto), then $\sigma: M \times M \to \mathbb{R}$ defined by $\sigma(x,y) = |g(x) - g(y)|$ is a metric on M.

To see this, observe that

(i) $0 \leq \sigma(x,y) < \infty$ for all pairs $x, y \in M$

(ii) $\sigma(x,y) = 0$ if and only if $|g(x) - g(y)| = 0$, if and only if $g(x) = g(y)$, which happens only when $x = y$ (why?)

$$\text{(iv)}\quad \sigma(x,y) = |g(x) - g(y)| = |g(x) - g(z) + g(z) - g(y)|$$

$$\leq |g(x) - g(z)| + |g(z) - g(y)| = \sigma(x,z) + \sigma(z,y)$$

Ex. (a) Define $d_1, d_2, d_3 : \mathbb{R}^2 \to [0,\infty)$ by $d_1(x,y) = |\tan^{-1}x - \tan^{-1}y|$, $d_2(x,y) = |x^3 - y^3|$, and $d_3(x,y) = |e^x - e^y|$ respectively. Then d_1, d_2, and d_3 are all metric functions on $\mathbb{R}$.

(b) Let $M = (0,\infty)$. Then $d_1, d_2, d_3 : (0,\infty)\times(0,\infty) \to [0,\infty)$ defined by $d_1(x,y) = |\sqrt{x} - \sqrt{y}|$, $d_2(x,y) = |\ln x - \ln y| = |\ln(\frac{x}{y})|$ and $d_3(x,y) = |\frac{1}{x} - \frac{1}{y}|$ respectively, are all metric functions on M.

Have you ever heard the beautiful voice of Anna German? If not, then you haven't really lived. In one of my favorite songs she sings " I will hear you a thousand miles accross. We are an echo of each other."

Notice that if she is at $x = 10$ and 'he' is at $y = 1010$, then under the metric $d(x,y) = |\frac{1}{x} - \frac{1}{y}|$, 1000 miles seems like $\frac{10}{101}$ or less than 0.1 mi.

Sometimes one must go far away to be close to another.

(c) Please note that a function can be a metric on one set and fail to be a metric on another. Take, for example, the function $d(x,y) = |x^2 - y^2|$. Then d defines a metric on $[0,\infty)$, but fails to be a metric on $\mathbb{R}$. (Why? Try to give more than one reason.)

We can expand our collection of metrics even further. To do this, we first prove the following lemma.

Lemma: Let $f: [0,\infty) \to [0,\infty)$ be any function with the following properties.

i) $f(x) = 0$ if and only if $x = 0$. Otherwise $f(x) > 0$

ii) f' is decreasing. That is, if $x < y$, then $f'(x) > f'(y)$

Then for any pair of points $x, y \in [0,\infty)$,

$f(x+y) \leq f(x) + f(y)$.

Proof: let $g(x) = f(x+y)$ and $p(x) = f(x) + f(y)$, where we regard y as a fixed number. We wish to show that $g(x) \leq p(x)$ or, equivalently, that $0 \leq p(x) - g(x)$. Notice that $\frac{d}{dx}(p(x) - g(x)) = p'(x) - g'(x) = f'(x) - f'(x+y) \geq 0$ by property (ii) of f. Thus, by the first derivative test,

$p(x)-g(x)$ is increasing for all $x\in(0,\infty)$, attaining its smallest value when $x=0$. Now $p(0)-g(0) = f(0)+f(y)-f(y) = f(y)-f(y) = 0$.

Thus $p(x)-g(x)\geq 0$ for all x and the desired result follows.

Thm: let $d: M\times M \to [0,\infty)$ be a metric function on M and suppose $f:[0,\infty)\to[0,\infty)$ satisfies properties (i) and (ii) of the above lemma. If $f'(t)>0$ for all $t\in(0,\infty)$, then $p: M\times M\to[0,\infty)$ given by $p(x,y)=f(d(x,y))$ defines another metric on M.

Proof: (a) Clearly $0\leq f(d(x,y))<\infty$ for all $x,y\in M$. Hence $0\leq p(x,y)<\infty$

(b) Suppose $p(x,y)=0$, then $f(d(x,y))=0$. By property (i) of f, this implies that $d(x,y)=0$, or $x=y$. Obviously $p(x,x)=0$

(c) Clearly $p(x,y)=p(y,x)$ for all $x,y\in M$

(d) $p(x,y) = f(d(x,y)) \leq f(d(x,z)+d(z,y)) \leq f(d(x,z))+f(d(z,y)) = p(x,z)+p(z,y)$ for all $x,y,z\in M$

Here the first inequality comes from the assumption that f is increasing and $d(x,y) \leq d(x,z) + d(z,y)$ and the second inequality is a consequence of the lemma just proved.

Since ρ satisfies properties (a)−(b), we see that ρ is a metric function.

Ex. (a) you should verify that $\rho(a,b) = \sqrt{|a-b|}$,
$\sigma(a,b) = \frac{|a-b|}{1+|a-b|}$, and $\tau(a,b) = \ln(|a-b|+1)$
each define metrics on $\mathbb{R}$.

(b) If d is any metric on M, verify that
$\rho(x,y) = \sqrt{d(x,y)}$, $\sigma(x,y) = \frac{d(x,y)}{1+d(x,y)}$, and $\tau(x,y) =$
$= \ln(d(x,y)+1)$ are also metrics on M.

Comprehension Check: Is $\rho(x,y) = \sqrt{\ln(|x^3-y^3|+1)}$ a metric function on $\mathbb{R}$?
How about $\sigma(x,y) = \frac{\sqrt{\ln(|x^3-y^3|+1)}}{1+\sqrt{\ln(|x^3-y^3|+1)}}$?